FUNCTIONAL GENOMICS OF NERVOUS SYSTEM DEVELOPMENT AND DISEASE

by

MICHAEL RYAN MILLER

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2011

DISSERTATION APPROVAL PAGE

Student: Michael Ryan Miller

Title: Functional Genomics of Nervous System Development and Disease

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Dr. John Postlethwait	Chair
Dr. Chris Doe	Advisor
Dr. Bruce Bowerman	Member
Dr. Patrick Phillips	Member
Dr. Tom Stevens	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2011

DISSERTATION ABSTRACT

Michael Ryan Miller

Doctor of Philosophy

Department of Biology

December 2011

Title: Functional Genomics of Nervous System Development and Disease

The goal of functional genomics is to elucidate the relationship between an organism's genotype and phenotype. A key characteristic of functional genomics is the use of genome-wide approaches as opposed to more traditional single-gene approaches. Genome-wide expression profiling is used to investigate the dynamic properties of transcriptomes, provides insights into how biological functions are encoded in genomes, and is an important technique in functional genomics. This dissertation describes the use of genome-wide expression profiling and other functional genomics techniques to address a variety of biological questions related to development and disease of the nervous system. Our results reveal novel and important insights into nervous system development and disease and demonstrate the power of functional genomics approaches for the study of nervous system biology. This dissertation also describes a novel technique called TU-tagging that facilitates cell type-specific RNA isolation from intact complex tissues. The isolation of RNA from specific cell types within a complex tissue is a major limiting factor in the application of genome-wide expression profiling, and TU-tagging can be used to address a wide array of interesting and important biological questions.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Michael Ryan Miller

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2011, University of Oregon
Bachelor of Science, Biology, 2006, University of Oregon

AREAS OF SPECIAL INTEREST:

Genetics and Genomics

PROFESSIONAL EXPERIENCE:

Graduate researcher, Department of Biology, University of Oregon, Eugene, 2006-2011

Teaching assistant, Department of Biology, University of Oregon, Eugene, 2006-2007

Undergraduate researcher, Department of Biology, University of Oregon, Eugene, 2004-2006

GRANTS, AWARDS, AND HONORS:

Aquaculture Workshop Student Travel Award, International Plant and Animal Genome Conference XIX, 2011

Pete von Hippel Graduate Education Endowment Award, University of Oregon Institute of Molecular Biology, 2010

Best Student Poster Presentation, Society for Developmental Biology Northwest Regional Meeting, 2009

Graduate Research Fellowship, National Science Foundation, 2008-2011

Integrative Graduate Education and Research Traineeship Program Associate Award, National Science Foundation, 2007-2010

Developmental Biology Traineeship, National Institutes of Health, 2007-2008

Graduate Teaching Fellowship, University of Oregon, 2006-2007

PUBLICATIONS:

Carney, T.D., Miller, M.R., Robinson, K.J., Bayraktar, O.A., Osterhout, J.A., Doe, C.Q. Functional genomics identifies neural stem cell sub-type expression profiles and genes regulating neuroblast homeostasis. Dev. Biol. Under review.

Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E. 3rd, Palti, Y., Doe, C.Q., Thorgaard, G.H. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. In press.

O'Rourke, S.M., Yochem, J., Connolly, A.A., Price, M.H., Carter, L., Lowry, J.B., Turnbull, D.W., Kamps-Hughes, N., Stiffler, N., Miller, M.R., Johnson, E.A., Bowerman, B. Rapid mapping and identification of mutations in Caenorhabditis elegans by RAD mapping and genomic interval pull-down sequencing. Genetics. In press.

Liu, C., Sage, J.C., Miller, M.R., Verhaak, R.G.W., Hippenmeyer, S., Vogel, H., Forman, O., Bronson, R.T., Nishiyama, A., Luo, L., Zong, H., 2011. Mosaic analysis with double markers reveals tumor cell of origin in glioma. Cell 146, 209-221.

Tran, K.D., Miller, M.R., Doe, C.Q., 2010. Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Development 137, 1421-1430.

Miller, M.R., Robinson, K.J., Cleary, M.D., Doe, C.Q., 2009. TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat. Methods 6, 439-441.

Yokoi, H., Yan, Y.L., Miller, M.R., BreMiller, R.A., Catchen, J.M., Johnson, E.A., Postlethwait, J.H., 2009. Expression profiling of zebrafish sox9 mutants reveals that Sox9 is required for retinal differentiation. Dev. Biol. 329, 1-15.

Lewis, Z.A., Shiver, A.L., Stiffler, N., Miller, M.R., Johnson, E.A., Selker, E.U., 2007. High density detection of restriction site associated DNA (RAD) markers for rapid mapping of mutated loci in Neurospora. Genetics 177, 1163-1171.

Miller, M.R., Atwood, T.S., Eames, B.F., Eberhart, J.K., Yan, Y.L., Postlethwait, J.H., Johnson, E.A., 2007. RAD marker microarrays enable rapid mapping of zebrafish mutations. Genome Biol. 8, R105.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17, 240-248.

Johnson, E.A., Liu, G., Miller, M.R., 2006. Methods of mapping polymorphisms and polymorphism microarrays. Patent Pub. No. WO/2006/122215.

ACKNOWLEDGMENTS

I thank B. Barut, S. Johnson, F. Pignoni, J. Rawls, J. Skeath, U. Walldorf, L. Zon, the Bloomington Stock Center, the Developmental Studies Hybridoma Bank, the Vienna Drosophila RNAi, and the Zebrafish International Resource Center for providing reagents; J. Boone, J. Dugas, J. Gleason, A. Henner, P.K. Loi, J. Murphy, A. Rapp, A. Selix, A. Starks, and M. Valle for technical support; B. Bowerman, C. Canestro, C. Doe, R. Galvao, T. Herman, W. Hong, M. Kohwi, S. Lai, Z. Lewis, S. O'Rourke, M. Raff, R. Read, B. Tasic, X. Wu, and Y. Zhu for critical comments on chapters of this dissertation; the European Molecular Biology Organization, Howard Hughes Medical Institute, Human Frontier Science Program Organization, National Center for Research Resources, National Institutes of Health, National Science Foundation, Pew Charitable Trusts, and Swiss National Science Foundation for financial support; my co-authors on chapters of this dissertation for the exciting and fruitful collaborations; and my dissertation advisory committee and other members of the University of Oregon Biology Department for invaluable advice and support throughout my tenure as a student.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

This dissertation includes previously published and unpublished co-authored material. Chapter II was previously published with co-authors Hayato Yokio, Yi-Lin Yan, Ruth BreMiller, Julian Catchen, Eric Johnson, and John Postlethwait. Chapter III is unpublished material with co-authors Travis Carney, Kristin Robinson, Omer Bayraktar, Jessica Osterhout, and Chris Doe. Chapter IV was previously published with co-authors Chong Liu, Johnathan Sage, Roel Verhaak, Simon Hippenmeyer, Hannes Vogel, Oded Foreman, Roderick Bronson, Akiko Nishiyama, Liqun Luo, and Hui Zong. Chapter V was previously published with co-authors Kristin Robinson, Michael Cleary, and Chris Doe.

The availability of complete genome sequences for many organisms has led to the emergence of a discipline called functional genomics. The overarching goal of functional genomics is to understand the biological function of sequence elements and in turn elucidate the relationship between an organism's genotype and phenotype. This lofty goal requires the development and application of global experimental approaches, and one characteristic of functional genomics is the use of genome-wide approaches as opposed to more traditional single-gene approaches. Genome-wide gene-expression profiling is one of the most important and widely-used techniques in functional genomics and is used to investigate the dynamic properties of cellular transcriptomes. This technique provides a wealth of data and can be used investigate diverse biological processes.

Chapters II, III, and IV of this dissertation describe the use of genome-wide expression profiling and other functional genomics techniques to address a variety of

biological questions related to development and disease of the nervous system. Chapter II describes an investigation of the mechanism of action of the Sox9 transcription factor which is mutated in humans with the severe disorder campomelic dysplasia. We used genome-wide expression profiling followed by other methods to compare wild-type zebrafish embryos to mutants lacking Sox9 activity and discovered novel roles for Sox9 in regulating Müller glia and photoreceptor cell numbers and organizing the neural retina. Chapter III describes an investigation of the genes regulating lineage pattern, tumor susceptibility, and homeostasis in neural stem cells (NSCs) of the *Drosophila* central nervous system (CNS). We used genome-wide expression profiling and found that the differential expression of surprisingly few genes can have profound effects on NSC lineage pattern and tumor susceptibility. We also performed an RNAi-based functional screen of genes identified with the expression profiling and identified many novel regulators of NSC homeostasis. Chapter IV describes an investigation of the cell type of origin of cancer in a mouse model of the deadly brain cancer malignant glioma. We used a variety of experiments, including genome-wide expression profiling, and found that oligodendrocyte precursor cells (OPCs) as opposed to other NSC-derived lineages or NSCs themselves were the glioma cell of origin. These three studies reveal novel and important insights into nervous system development and disease and demonstrate the power of functional genomics approaches for the study of nervous system biology.

Chapter V of this dissertation describes a novel technique called TU-tagging that facilitates cell type-specific RNA isolation from intact complex tissues. We demonstrated that a combination of spatially restricted uracil phosphoribosyltransferase (UPRT) expression with 4-thiouracil delivery can be used to label and purify RNA from specific

cell types within the *Drosophila* CNS. In this technique, 4-thiouracil is modified and subsequently incorporated into newly synthesized RNA only in cells expressing UPRT. Thio-substituted nucleotides are not a natural component of nucleic acids, and the resulting thio-labeled RNA can be readily tagged and purified. Therefore, even if RNA is isolated from the whole organism, RNA from the cells expressing UPRT can be recovered by purifying labeled RNA. This method is useful for isolating RNA from cell types that are difficult to isolate by dissection or dissociation methods and should work in many organisms, including mammals and other vertebrates. The isolation of RNA from specific cell types within a complex tissue is a major limiting factor in the application of genome-wide expression profiling, and TU-tagging can be used to address a wide array of interesting and important biological questions.

CHAPTER II

EXPRESSION PROFILING OF ZEBRAFISH SOX9 MUTANTS REVEALS THAT SOX9 IS REQUIRED FOR RETINAL DIFFERENTIATION

This work was published in volume 329 of the journal *Developmenal Biology* in May 2009. Hayato Yokio and myself performed the microarray-based expression analysis. Yi-Lin Yan, Hayato Yokio, and Ruth BreMiller performed the in situ hybridization and immunohistochemistry analysis. Julian Catchen performed the phylogenetic analysis. John Postlethwait and Eric Johnson were the principle investigators for this work.

Introduction

Haploinsufficiency for *SOX9* results in the human disease campomelic dysplasia (CD), a syndrome characterized by skeletal abnormalities and male-to-female sex reversal (Foster et al., 1994; Wagner et al., 1994). This clinical phenotype shows that *Sox9* function is important for skeletal development and male sex differentiation, and these processes have been foci of *Sox9* research (Bi et al., 1999; Marshall and Harley, 2000; Vidal et al., 2001). CD patients show various other clinical features, including micrognathia, cleft palate, hypoplastic lungs, defects in olfactory tissues, and heart and renal malformations (Houston et al., 1983). Most patients die in early infancy because of respiratory problems, and some survivors show hearing loss and myopia (Mansour et al., 2002). These descriptions identify various essential functions of *SOX9*, for example, in neural crest cells (NCC), heart, lung, ear, and other tissues. Intensive analyses of *Sox9* using conditional knockout mice and other model organisms have also assigned functions to *Sox9*, including the differentiation of neural crest cells (Spokony et al., 2002; Mori-

Akiyama et al., 2003; Saint-Germain et al., 2004; Cheung et al., 2005; Sakai et al., 2006), glial cells in the spinal cord (Stolt et al., 2003), paneth cells in the intestine (Bastide et al., 2007; Mori-Akiyama et al., 2007), notochord (Barrionuevo et al., 2006), heart valves (Akiyama et al., 2004; Lincoln et al., 2007), pancreas (Seymour et al., 2007), otic vesicle (Barrionuevo et al., 2008) and gonad (Chaboissier et al., 2004). *Sox9* is now recognized as a multi-functional gene that plays essential roles in various tissues during vertebrate development.

Sox9 encodes a transcription factor of the *Sry*-related HMG box (*Sox*) family that binds to cis-regulatory DNA elements to control the transcription of down-stream target genes. Genes shown to be regulated by *Sox9* include *Col2a1* and *Col11a2* in the skeleton and *Amh*/*Mis* in the gonad (Bell et al., 1997; Lefebvre et al., 1997; Ng et al., 1997; Bridgewater et al., 1998; De Santa Barbara et al., 1998; Arango et al., 1999; Bi et al., 1999). Recent microarray analysis using cultured human cells revealed that the calcium binding proteins S100A1 and S100B are transcriptional targets of SOX9 and its coactivators SOX5 and SOX6 (Saito et al., 2007). Given the multi-functionality of *Sox9*, however, most of its targets remain unidentified. Identification of *Sox9* target genes will improve our understanding of the biological mechanisms by which *Sox9* functions to regulate development.

Zebrafish has two copies of *Sox9*, called *sox9a* and *sox9b* (Chiang et al., 2001). These co-orthologs of mammalian *Sox9* resulted from a whole genome duplication event that occurred after teleost lineage segregated from the tetrapod lineage and before the diversification of teleost fish (Amores et al., 1998; Postlethwait et al., 1998; Taylor et al., 2003; Jaillon et al., 2004). After the genome duplication event, duplicate gene copies

often evolved in a paralog-specific manner involving, at least in part, subfunction partitioning (Force et al., 1999; Postlethwait et al., 2004). The expression patterns of *sox9a* and *sox9b* overlap in some regions and are gene-specific in other domains, and the sum of their expression patterns is similar to the mouse *Sox9* expression pattern (Chiang et al., 2001; Cresko et al., 2003; Yan et al., 2005).

The zebrafish mutation *jellyfish* (*jef*), which disrupts *sox9a*, causes skeletal defects due to abnormal cartilage formation (Yan et al., 2002). The zebrafish *sox9b* deletion mutant shows reduction of cartilage, and the sox9a;sox9b double mutant shows more severe defects revealing synergistic, additive and redundant functions of the two zebrafish co-orthologs of *Sox9* (Yan et al., 2005). The partitioning of *Sox9* subfunctions between the two zebrafish co-orthologs facilitates the analysis of *Sox9* functions that were difficult to analyze in mammals due to pleiotropy (Yan et al., 2005). In mouse, heterozygous mutants (*Sox9*+/−) showed skeletal abnormalities and cleft palate, and die perinatally because of haploinsufficiency, as do most human CD patients (Bi et al., 2001); thus, homozygous mutant mouse embryos are not available. Fortunately, tissue-specific conditional knockout systems allow the investigation of *Sox9* function in specific tissues (Kist et al., 2002; Chaboissier et al., 2004). But if *Sox9* acts in a tissue that has not been examined by tissue-specific knockouts, we will not learn the full role of *Sox9* in development. Making use of the zebrafish *sox9a* and *sox9b* mutants, whose heterozygotes are viable and fertile, we performed microarray analysis to compare the expression profile between homozygous mutant and wild-type embryos. Microarray is a powerful technique that enables us to conduct a genome wide screening of changes in gene expression between different biological groups, however at the same time, any detected

changes in gene expression may be due to secondary effects, including indirect non-cell-autonomous effects.

We evaluated potential *Sox9* targets identified by microarray by comparing the expression patterns of putative targets with the expression patterns of *sox9a* and *sox9b*. We also tested for reduced expression of putative targets in *sox9* mutant embryos by in situ hybridization. Validated downstream candidate targets of *sox9* included previously known targets such as *collagen type II alpha-1a* (*col2a1a*) and *collagen type XI alpha-2* (*col11a2*) (Bell et al., 1997; Lefebvre et al., 1997; Ng et al., 1997; Bridgewater et al., 1998; Bi et al., 1999), and novel targets, including *cone rod homeobox* (*crx*), *retinoschisis 1* (*rs1*), *calbindin 2* (*calb2*) genes and other genes expressed in developing retina. The demonstration that *sox9b* mutant embryos have eye defects supports the role of *sox9* candidate targets. *Sox9* is expressed in the retina of developing mouse embryos (Ihanamäki et al., 2002), but its function in the vertebrate retina has not been characterized. Thus, microarray analysis using zebrafish mutants coupled to mutant phenotypes, uncovered a previously unidentified function of *Sox9*. We also present here the detailed analyses of eye defects observed in the *sox9b* mutants and show the importance of *Sox9* in retinal differentiation.

Methods

Zebrafish

Zebrafish were maintained under standard conditions (Westerfield, 2000) and embryos were staged as described (Kimmel et al., 1995). The *sox9a* mutation *jef*^{tw37}, the *sox9b* deletion mutant *b971*, and the *sox9a*^{tw37};*sox9b*^{b971} double mutant stock have been

described previously (Piotrowski et al., 1996; Yan et al., 2002; Yan et al., 2005). Embryos were collected from matings of *sox9a*;*sox9b* double heterozygotes. At 2 days post fertilization (dpf), *sox9b* mutant embryos are distinguished from wild-type embryos by a curly tail phenotype, and *sox9a*;*sox9b* double mutant embryos are distinguished by lack of otic vesicles, while *sox9a* mutant embryos are morphologically indistinguishable from wild-type embryos until 3 dpf. All work with vertebrate animals was approved by the University of Oregon Institutional Animal Care and Use Committee.

RNA extraction

Total RNA was extracted from embryos using TRI reagent. RNA quantity was measured using NanoDrop ND-1000 spectrophotometer. Approximately 1 μg of total RNA was extracted from individual embryos at 2 dpf.

Microarray analysis

Zebrafish DNA microarray containing 16,399 65-mer oligonucleotides representing 16,228 gene clusters and 171 control features was kindly provided by Steve Johnson and John Rawls (Rawls et al., 2004). The gene clusters arise from various developmental stages and organs, including maternal RNAs, shield, early gastrulation and segmentation stage, and various adult tissues and organs, including brain, heart, kidney, retina, ovary and testis.

For each experiment, 20 μg of total RNA from wild-type or mutant embryos was used to synthesize cDNA labeled with Cy3 or Cy5-dUTP using SuperScript III reverse transcriptase. The Cy3 labeled cDNA was mixed with Cy5 labeled cDNA and purified with QIAquick PCR purification kit, dried in a speed vac and resuspended in hybridization buffer (50% formamide, 5 × SSC, 1% SDS, 1 mg/ml calf thymus DNA).

After denaturing, labeled cDNA was applied onto the microarray and covered with a coverslip, incubated in a hybridization chamber, and hybridized in a 42 °C water bath overnight. Following the hybridization, the microarray was washed with an SSC series and dried by centrifugation. The microarray was immediately scanned using Genepix 4000B scanner. Three sets of experiments with dye-swap, in total six hybridizations, were conducted using independently isolated RNA samples from three different sets of animals as biological replicates. Each microarray was scanned with two or three different laser intensities, depending on the variation in signal intensities, to recover data of wider range of gene expression.

GenePix Pro 3.0 image analysis software was used to measure fluorescence signal intensities of the array elements. Ratios were normalized to the overall difference in signal intensities from each channel. Microarray data were transferred to Microsoft Excel, which was also used for calculations and data sorting. The “Median of Ratios” (MR, the median of Cy3 and Cy5 ratios taken across the array element) was used for the analyses. Each microarray has two or three scan results (MRs) obtained with different laser intensities, from which the “Median of the MR” (MeMR) and the “Maximum of the MR” (MaMR) were calculated. Data from each hybridization were merged and sorted according to the Median of the MeMR. The probability of rejecting the hypothesis that gene expression was not changed between wild-type and mutant ($\log2 = 0$) was calculated by two tailed Student's t-test. Candidate genes were picked from the list of the summary result of the microarray. The results from each single experiment (MeMR and MaMR) and scanned images were also considered. Genes that showed inconsistent results in the dye-swap experiment were removed from further analyses.

Cloning of candidate genes and whole-mount in situ hybridization

Candidate genes were cloned by RT-PCR, except for *klf2b*, which was kindly provided by Bruce Barut and Len Zon (Oates et al., 2001), and *calb2a* and *calb2b*, which were obtained from the I.M.A.G.E. Consortium EST collection. PCR primers were designed based on previous descriptions or genome information from the Ensembl database (Flicek et al., 2008). PCR products were cloned into pCR4-TOPO vector. DIG-labeled RNA probe was synthesized with T3 or T7 RNA polymerase. Whole-mount in situ hybridization was performed as described in Yan et al. (2005).

Phylogenetic analysis

Amino acid sequences were aligned using CLUSTAL X (Thompson et al., 1997) and the alignment was further checked by eye using Seaview software (Galtier et al., 1996). A neighbor-joining (NJ) tree (Saitou and Nei, 1987) was plotted using NJplot (Perrière and Gouy, 1996).

Sectioning of the whole-mount in situ samples and in situ hybridization on sections

Samples from whole-mount in situ hybridization were embedded in paraffin and sectioned in the Histology Facility of the University of Oregon. In situ hybridization on cryosections were performed as described (Rodríguez-Marí et al., 2005).

Immunohistochemistry and confocal observation

Retinal Müller glial cells were labeled by anti-carbonic anhydrase antibody 1:200 (Jensen et al., 2001), the photoreceptor layer was labeling using zpr-3 antibody 1:200 from the Zebrafish International Resource Center (ZIRC) (Jensen et al., 2001), and amacrine cells were labeled by anti-GABA antibody 1:1000 (Sandell et al., 1994). Nuclear layers of the retina were stained by the nuclear dye TO-PRO-3 iodide.

Fluorescent images were analyzed using Bio-Rad Radiance 2100MP confocal microscope system.

Results

Identification of Sox9 down-stream targets

To identify genes under control of *Sox9* during vertebrate embryogenesis, we compared the gene expression profile of wild-type and *sox9a*;*sox9b* double mutant zebrafish embryos at 2 dpf using a zebrafish oligonucleotide microarray (Rawls et al., 2004). The 2 dpf time point is late enough so that double mutant embryos are morphologically distinguishable from wild-type and single mutant embryos, but early enough so that most secondary effects will not have yet taken place; nevertheless, some secondary downstream genes may also be detected because *sox9* genes express as early as 1 dpf (Chiang et al., 2001). The microarray represents 16,228 gene clusters from EST collections of various embryonic stages and a variety of adult organs, including brain, eye, heart, kidney, liver, testis and ovary. To minimize erroneous expression results, three sets of experiments with dye-swap were conducted using independently isolated RNA samples. Data were analyzed and sorted with GenePix Pro software and Microsoft Excel, and candidates were picked according to the criteria described in Methods. Because Sox9 is a transcription activator (Südbeck et al., 1996), we focused on genes that are down-regulated in mutant embryos.

The *b971* mutant allele of *sox9b* used in this study deletes *sox9b* and some surrounding genes including *sox8* (Yan et al., 2005). Because the morpholino knock down of *sox9b* phenocopies the $sox9^{b971}$ defects, the phenotype observed in $sox9^{b971}$ at

this developmental stage is due solely to loss of *sox9b* function (Yan et al., 2005). The *sox8* gene is not expressed at this time in development (Yan et al., 2005). The Ensembl zebrafish genome database (Flicek et al., 2008) was used to evaluate the genomic position of candidate genes. Although *pvalb1* (*parvalbumin 1*) and *pvalb4*, *pvalb8* consistently appeared as down-regulated genes, they are removed by the *sox9*^{b971} deletion, and so are excluded from the candidates.

To test whether candidates derived from the microarray analysis were valid, we utilized in situ hybridization in wild-type and mutant embryos to evaluate potential *Sox9* targets. We cloned the following candidate genes: *calbindin 2a* (*calb2a*), *calbindin 2b* (*calb2b*), *collagen type II alpha-1a* (*col2a1a*), *collagen type XI alpha-2* (*col11a2*), *cone rod homeobox* (*crx*), *crystallin alpha A* (*cryaa*), *crystallin beta B1* (*crybb1*), *crystallin gamma M2c* (*crygm2c*), *kruppel like factor 2b* (*klf2b*), *neurod*, *retinoschisis (X-linked, juvenile) 1* (*rs1*), *sox4a*, *syndecan 2*, *visual system homeobox 1* (*vsx1*) and unknown ESTs, AI585100, AI722369, AI942960 and AW777642. The expression of candidate genes was compared among wild-type, *sox9a* mutant, *sox9b* mutant and *sox9a*;*sox9b* double mutant embryos at 2 dpf, the same stage used for the microarray analysis. We describe below genes identified in the microarray experiments whose expression change was verified in *sox9* mutant embryos. Candidate genes could be categorized into two major groups: skeletogenesis and retinal development (Table 1; see Appendix A for all tables).

col11a2 expression is regulated by both zebrafish Sox9 co-orthologs

The collagen genes *col2a1a* and *col11a2* showed 3.148 and 1.983 times higher expression in wild-type compared with double mutants, respectively, and dye-swap

experiments confirmed results. Because their mouse orthologs are known as direct targets of *Sox9* (Bell et al., 1997; Lefebvre et al., 1997; Ng et al., 1997; Bridgewater et al., 1998; Bi et al., 1999; Liu et al., 2000), and because *col2a1a* expression is known to be reduced in *sox9* mutants (Yan et al., 2005), these results validate our experimental strategy. Here, we tested whether *col11a2* is a target of *sox9a* and/or *sox9b* by analyzing the expression of *col11a2* in wild-type and *sox9* mutants.

In situ hybridization experiments showed that at 2 dpf, *col11a2* was expressed in neurocranial cartilages (eg. ethmoid plate), pharyngeal cartilages (eg, mandibular cartilage), otic vesicle, and pectoral fin bud (Figs. 1A, E; see Appendix B for all figures). In *sox9a* mutants, expression of *col11a2* disappeared from the neurocranium and pharyngeal arches, and was reduced in the opercle (Figs. 1B, F). This expression pattern was consistent with the previously described *sox9a* mutant phenotype in which all neurocranial cartilage elements and most pharyngeal arch cartilages were missing from *sox9a* mutants; dermal bone elements, however, were present but small in *sox9a* mutant embryos (Piotrowski et al., 1996; Yan et al., 2002; Yan et al., 2005). The expression of *col11a2* was also affected in *sox9b* mutants, but in a fashion different than in *sox9a* mutants: *col11a2* expression was greatly reduced in the mandibular and ceratohyal cartilages in *sox9b* mutants, but was retained in the ethmoid plate and trabeculae of the neurocranium (Figs. 1C, G), consistent with the *sox9b* mutant phenotype (Yan et al., 2005). In *sox9a*;*sox9b* double mutants, expression of *col11a2* disappeared in all neurocranial and pharyngeal cartilages. Because the otic vesicle is usually missing in double mutants (Yan et al., 2005), *col11a2* expression in the otic vesicle vanished in *sox9a*;*sox9b* double mutant embryos. Expression of *col11a2* in the pectoral fin bud was

greatly reduced in the *sox9a*;*sox9b* double mutant (Figs. 1D, H).

Expression patterns of the two zebrafish *sox9* genes can help explain the observed phenotypic differences between the *sox9a* and *sox9b* mutants: at 2 dpf, *sox9a* is expressed in the ethmoid plate, mandibular and ceratohyal (Fig. 1I), whereas *sox9b* is expressed in the mandibular and ceratohyal cartilages, but not in the ethmoid plate (Fig. 1J). These expression patterns are consistent with the *col11a2* expression phenotype in the mutants: *col11a2* expression in the neurocranium depends on *sox9a* function, but expression in the pharyngeal cartilages depends on both *sox9a* and *sox9b*. We conclude that both *sox9a* and *sox9b* regulate *col11a2* in a tissue-specific fashion related to the *sox9* genes' expression domains. Thus, the regulatory relationship between *Sox9* and *Col11a2* is conserved from teleost fish to mammals.

Regulation of retinal genes by Sox9

A group of genes expressed in the eye was down regulated in mutant embryos at 2 dpf. Because the size of the eye was reduced in *sox9b* and *sox9a*;*sox9b* double mutant embryos compared to wild-type and *sox9a* mutant embryos, we sought to test the hypothesis that *sox9b* is required for eye development.

In situ hybridization experiments showed broad expression of *crx* throughout wild-type retinas at 2 dpf (Fig. 2A). The expression of *crx* was severely affected in retinas of *sox9b* mutants and *sox9a*;*sox9b* double mutant embryos (Figs. 2B, C), except for a small patch of cells in the ventral part of the retina (Figs. 2B, C). *Crx* is a member of the *otd*/*Otx* family of *paired-like* homeobox genes and is essential for retinal photoreceptor cell differentiation and maintenance in mouse (Furukawa et al., 1997; Furukawa et al., 1999). Zebrafish *crx* is expressed in the epiphysis and in proliferating retinal progenitors

(Liu et al., 2001) and may function in the differentiation of photoreceptors and other cells of the retina (Shen and Raymond, 2004).

In wild-type embryos (Fig. 2D), *rs1* was expressed strongly throughout the entire retina (Thisse and Thisse, 2004). In contrast, in *sox9b* mutants and double mutants (Figs. 2E, F), *rs1* expression was retained only in the ventral part of the retina in the same domain as found previously for *crx* (Figs. 2E, F). Rs1 is a secreted protein with putative functions in cell adhesion and/or signaling (Wu et al., 2005; Molday et al., 2007). Mutations in the human *RS1* gene cause X-linked retinoschisis, a disease involving intraretinal splitting due to degeneration (Sauer et al., 1997). Zebrafish *rs1* was cloned in the Zebrafish Gene Collection Project (ZGC), and the ZFIN database reported its expression pattern to include a few cells in branchial arches, epiphysis, and dorsal telencephalon in addition to the retina (Thisse and Thisse, 2004).

Consistent with previous work (Passini et al., 1997; Passini et al., 1998), we found expression of *vsx1* in the developing retina of wild-type embryos (Fig. 2G). In the retina of *sox9b* and *sox9a*;*sox9b* double mutant embryos, expression of *vsx1* was down-regulated, with the dorsal and caudal portions of the eye most severely reduced (Figs. 2H, I). Vsx1 is a *paired-like* homeobox protein, first isolated from goldfish retina (Levine et al., 1997). Gene knockout analyses in mouse showed that *Vsx1* is required for retinal cone bipolar cell differentiation (Chow et al., 2004; Ohtoshi et al., 2004).

Our in situ hybridization analyses (Fig. 2J) confirmed that *neurod* is expressed in both outer and inner nuclear layers of the retina and in the telencephalon in wild-type embryos (Korzh et al., 1998; Mueller and Wullimann, 2002; Rauch et al., 2003; Ochocinska and Hitchcock, 2007). Expression of *neurod* in the retina was severely

reduced in *sox9b* and *sox9a*;*sox9b* double mutant embryos (Figs. 2K, L). Because the expression of *neurod* was normal in the telencephalon, we conclude that *sox9b* is important for *neurod* expression in the retina but not in other parts of the nervous system. NeuroD is a basic helix–loop–helix (bHLH) transcription factor that can convert epidermal cells into neurons (Lee et al., 1995) and activate insulin gene transcription (Naya et al., 1995; Chae et al., 2004).

Our results confirmed expression of *sox4a* in the retina and developing brain of wild-type embryos (Fig. 2M) (Rauch et al., 2003; Thisse and Thisse, 2004), but showed that *sox4a* was not expressed in the retina of *sox9b* mutants and *sox9a*;*sox9b* double mutants; in contrast to the retina, expression of *sox4a* in the telencephalon and diencephalon was not affected in mutant animals (Figs. 2N, O). Sox4 is a *Sox* family transcription factor belonging to the *SoxC* group (*Sox4*, *-11*, and *-12*), while *Sox8*, *-9*, and *-10* are in the *SoxE* group (Bowles et al., 2000). *Sox4* is required for development of lymphocytes, heart and pancreas (Schilham et al., 1996; Ya et al., 1998; Wilson et al., 2005). Zebrafish has two co-orthologs of *Sox4* called *sox4a* and *sox4b*, that appear to have evolved in a duplication event at the base of the teleost radiation (Mavropoulos et al., 2005). In zebrafish, *sox4a* is widely expressed in the central nervous system including telencephalon, diencephalon, midbrain, rhombomeres, and retina (Rauch et al., 2003; Thisse and Thisse, 2004).

These expression results show that the full retinal expression of *crx*, *rs1*, *vsx1*, *neurod* and *sox4a* depends on *sox9b* but not *sox9a* function. This conclusion is consistent with the expression pattern of zebrafish *sox9a* and *sox9b*: at 2 dpf, *sox9b* is expressed in the ciliary marginal zone of the retina (Figs. 2S, T) but *sox9a* is not (Figs. 2P, Q) (Chiang

et al., 2001; Cresko et al., 2003; Yan et al., 2005).

Co-orthologs of Calb2 are down-regulated in the sox9b mutant retina

The calcium binding proteins *calbindin 2a* (*calb2a*, previously known as *calb2* or *calretinin*) and its closely related paralog *calbindin 2b* (*calb2b*, previously known as *calbindin 2-like*) were down-regulated in s*ox9b* and *sox9a*;*sox9b* double mutant embryos. These two genes have not previously been well characterized in zebrafish. Our phylogenetic analysis showed that *Calb2a* and *Calb2b* proteins branch as sisters of the tetrapod *Calb2* clade with high bootstrap support (Fig. 3A).

To evaluate the evolutionary origin of the zebrafish *calb2* genes, we analyzed conserved synteny around *calb2a* and *calb2b* using our orthology and conserved synteny database based on reciprocal best blast hit (RBH) analysis (Catchen et al., 2008). Red dots in Fig. 3B represent the chromosomal positions of zebrafish orthologs of a 20 Mb segment of human chromosome 16 (Hsa16) surrounding the *CALB2* gene. Orthologous chromosome segments appear as clusters of red dots. This analysis showed that a chromosomal segment of zebrafish chromosome 7 (Dre7) around *calb2a* and a segment of Dre18 around *calb2b* correspond to duplicated copies of the region surrounding *CALB2* in Hsa16 (Fig. 3B). These results show that *calb2a* and *calb2b* are co-orthologs of tetrapod *Calb2* and likely arose in the genome duplication that preceded the teleost radiation (Amores et al., 1998; Postlethwait et al., 1998; Jaillon et al., 2004).

Our expression analysis showed that in wild-type embryos, *calb2a* was expressed in the retina and in neurons of the cranial ganglia and in rhombomeres of the hindbrain (Fig. 3C), confirming directly submitted data in ZFIN (Rauch et al., 2003; Thisse and Thisse, 2004). The expression of *calb2b* appeared in the ganglion cell layer of the retina,

in the olfactory placode, tegmentum, midbrain-hindbrain boundary (MHB) and ventral hindbrain (Fig. 3F), which confirms data in ZFIN (Rauch et al., 2003; Thisse and Thisse, 2004). Both *calb2* genes were expressed in developing retina, but their expression patterns differed: *calb2a* was expressed widely in the retina including the ganglion cell layer, but *calb2b* was expressed only in the ganglion cell layer of the retina (Figs. 3C, F). In other parts of the nervous system, *calb2a* was expressed in cranial ganglia and rhombomeres of the hindbrain, whereas *calb2b* was expressed in the olfactory placode, tegmentum, MHB and ventral hindbrain. In mouse, high-throughput gene expression analyses by RNA in situ hybridization showed that *Calb2* is expressed in developing retina, olfactory bulb, tegmentum and hindbrain (Visel et al., 2004). Although description of the *Calb2* expression pattern in mouse is not detailed enough to know its cellular distribution within the retina, this domain appears to be ancient and conserved in mouse *Calb2* and partitioned into zebrafish co-orthologs *calb2a* and *calb2b*.

Expression analysis in *sox9* mutant embryos verified our microarray results. In *sox9b* mutants and in *sox9a*;*sox9b* double mutants, *calb2a* expression did not appear in the eye, although other *calb2a* expression domains, including the cranial ganglia and rhombomeres of the hindbrain, were not affected (Figs. 3D, E). Likewise, in *sox9b* mutant embryos and in *sox9a*;*sox9b* double mutant embryos, *calb2b* expression was missing from the retina, although other expression domains were not affected (Figs. 3G, H).

These results show that *sox9b* regulates expression of both *calb2a* and *calb2b* in the developing retina. As mentioned above, *calb2a* and *calb2b* are teleost fish co-orthologs of the tetrapod *Calb2* gene. Although the expression patterns of *calb2a* and

calb2b have diverged, expression of both genes in the retina depends on *sox9b*, suggesting that this relationship likely existed in the pre-duplication ancestral gene.

sox9b controls genes expressed in several retinal layers

As Fig. 2 and Fig. 3 show, a number of genes expressed in the retina were down-regulated in *sox9b* mutants. To learn which specific retinal cell types are regulated by Sox9, we investigated detailed expression domains in tissue sections. In wild-type embryos, *vsx1* was expressed in the outer part of the inner nuclear layer (Fig. 4A), but in *sox9b* mutants and in *sox9a*;*sox9b* double mutants, the expression domain was reduced dorsally (Figs. 4B, C). Expression of *rs1* was observed in the outer part of the inner nuclear layer and the outer nuclear layer in the wild-type retina (Fig. 4D), while in mutant embryos, *rs1* expression was reduced (Fig. 4E) or not detected (Fig. 4F). Expression of *calb2a* was observed in the inner nuclear layer (Fig. 4G) and *calb2b* was expressed in the ganglion cell layer (Fig. 4J). These expression domains were both severely reduced in mutant embryos with a small portion of ventral expression occasionally remaining (Figs. 4H, I, K, L). These results show that *sox9b* plays important roles in several different retinal cell layers, including the ganglion cell layer, inner nuclear layer and outer nuclear layer.

Retinal differentiation in zebrafish sox9 mutants in later development

Expression profiling of wild-type and *sox9a*;*sox9b* double mutants identified seven genes expressed in the developing retina as downstream of *Sox9* and hence potentially early *Sox9* targets. As shown previously, *sox9b* is expressed in the developing eye by 24 hpf (Chiang et al., 2001; Cresko et al., 2003; Yan et al., 2005), and continues to be expressed in the retina at 2 dpf (Figs. 2S, T), so it is reasonable to conclude that *Sox9*

functions in retinal differentiation by regulating down-stream target genes in the zebrafish embryo.

To learn how mutant phenotypes change after 2 dpf, we examined the expression of putative *sox9* targets in 3 dpf embryos. At 3 dpf, *calb2a* was expressed in the ganglion cell layer and inner nuclear layer (Fig. 5Aa), and *calb2b* was expressed in the ganglion cell layer and in certain cells in the inner nuclear layer of wild-type embryos (Fig. 5Ba). The expression of neither *calb2a* nor *calb2b* was affected in *sox9a* mutants, but expression of both genes was reduced in the dorsal and ventral retina in *sox9b* mutants and in *sox9a*;*sox9b* double mutants (Figs. 5Ac, Ad, Bc, Bd). The outer and inner nuclear layers both expressed *crx* (Fig. 5Ca). In *sox9b* mutants and *sox9a*;*sox9b* double mutants, the same layers expressed *crx*, but expression was reduced in the most dorsal and most ventral expression domains (Figs. 5Cc, Cd). Expression of *neurod* was observed in the outer nuclear layer and in a few cells in the inner nuclear layer (Fig. 5Da), as previously reported (Ochocinska and Hitchcock, 2007). Expression of *neurod* was not affected in *sox9a* mutants (Fig. 5Db), but expression of *neurod* in the outer nuclear layer was less clear in *sox9b* and *sox9a*;*sox9b* double mutants, perhaps because retinal layers did not form properly in these mutants (Figs. 5Dc, Dd). The outer and inner nuclear layers both expresed *rs1*, except for the ciliary marginal zone (Fig. 5Ea). The expression was reduced only in the dorsal-most and ventral-most expression domains in *sox9b* mutants and in *sox9a*;*sox9b* double mutants (Figs. 5Ec, Ed). Expression of *sox4a* appeared in a proportion of inner nuclear layer cells and in some cells in the dorsal and ventral parts of the ganglion cell layer in wild types and *sox9a* mutants (Figs. 5Fa, Fb). In *sox9b* and double mutants, however, *sox4a* expression in the inner nuclear layer appeared to be a bit

stronger than in the wild-type retina (Figs. 5Fc, Fd), suggesting an expansion of the *sox4a+* cell type. Sections showed that *vsx1* was expressed in the inner nuclear layer (Fig. 5Ga), but expression in the dorsal and ventral parts of the retina was reduced in *sox9b* and in *sox9a*;*sox9b* double mutants (Figs. 5Gc, Gd).

Expression of *col2a1a*, a known target of Sox9 in cartilaginous tissue, appeared in the iris (Fig. 5Ha). In contrast to cartilaginous tissue, *col2a1a* expression in the iris was retained in *sox9a* mutants, *sox9b* mutants, and in *sox9a*;*sox9b* double mutant embryos (Figs. 5Hb, Hc, Hd). These results showed that the regulatory relationship of Sox9 → *Col2a1* that is well established for cartilage does not hold for the iris.

Taken together, expression analysis on tissue sections suggest that *sox9b*, but not *sox9a*, is required for retinal development. The requirement for *Sox9* is most evident around 2 dpf and becomes milder at 3 dpf in *sox9b* mutants and in *sox9a*;*sox9b* double mutants, suggesting that other gene(s) can partially compensate for the lack of *sox9b* function on the expression of the candidate genes at 3 dpf. Because candidate genes expressed in or near the ciliary marginal zone were sensitive to loss of *sox9b* function at 3 dpf, *sox9b* may function in cell proliferation and/or retinal layer organization.

Independent of the microarray analysis, we had noticed that *sox9b* mutants have smaller eyes than wild types. We investigated eye development in wild-types and in *sox9a*, *sox9b* and double mutants to elucidate the function of *Sox9* in retinal differentiation in larval fish. By 5 dpf, most cell types in the retina have already differentiated (Fadool and Dowling, 2008). Antibody staining for specific cell layers revealed that cells expressing the Müller glial cell marker carbonic anhydrase (Figs. 6A, E, I, M) (Jensen et al., 2001) and the photoreceptor cell marker zpr-3 (Figs. 6B, F, J, N)

(Jensen et al., 2001) are reduced in number and disorganized in the retinas of *sox9b* mutants and *sox9a*;*sox9b* double mutants (Figs. 6I, J, M, N). In contrast, *sox9a* mutant retinas had no obvious defects, even though *sox9a* is expressed in the inner nuclear layer at 68 hpf (Fig. 2R). These results on 5dpf animals show that *sox9b* is required for differentiation of the Müller glial cell layer and photoreceptor cell layer in zebrafish. Glial cells are required for proper migration of neurons, so the disorganized retina might be secondary due to a primary defect in Müller glial cells. *Sox9* is involved in glial cell development in mouse spinal cord (Stolt et al., 2003), and here we showed that *sox9b* is required for retinal Müller glial cells. Although different glial types are involved, the role of *Sox9* in glial cell development may be conserved between Müller glia and glia of the central nervous system.

Expression of klf2b and EST AI722369 in the pectoral fin bud disappeared in sox9 mutants

The microarray experiments showed that *klf2b* was down regulated 2.344 fold in double mutants. Expression of *klf2b* was observed in the central mesenchyme of the pectoral fin bud and in the adjacent cleithrum of the shoulder girdle (Figs. 7A, C), as previously described (Oates et al., 2001; Thisse et al., 2001). In double *sox9* mutants, expression of *klf2b* in the fin bud disappeared but expression in the cleithrum was not affected (Figs. 7B, D). Consistent with these defects, *sox9a* and *sox9b* were expressed in the pectoral fin bud, but the cleithrum expressed neither *sox9a* nor *sox9b* (Yan et al., 2005). This result is consistent with the appendage defect observed in the *sox9* mutant pectoral fin: only the cleithrum was retained while other fin structures were greatly reduced in *sox9a*;*sox9b* double mutants (Yan et al., 2005).

An EST of unknown function (AI722369) was 5.673 times down regulated in double mutant embryos. The AI722369 nucleotide sequence has more than 98% identity over about 500 nucleotides to sequences assigned to at least five different chromosomes in the Ensembl genome database, suggesting that related sequences are dispersed in the zebrafish genome. AI722369 is similar to the uncharacterized human hypothetical protein LOC100131068. Despite appearing to be a repeated element, AI722369 was expressed in a strongly specific manner in time and space: in the pectoral fin bud in two parallel domains in the position of presumptive pectoral fin muscle (Fig. 7E). Furthermore, AI722369 expression disappeared from *sox9a*;*sox9b* double mutants (Fig. 7F), suggesting that it is downstream of *Sox9* activity. In the wild-type pectoral apparatus, *sox9a* is expressed strongly in the scapulocoracoid of the pectoral girldle but was weaker in the blade of the fin bud and did not appear in the cleithrum of the pectoral girdle. Reciprocally, *sox9b* is expressed strongly in the chondrogenic core of the fin bud but not in the scapulocoracoid or the cleithrum of the pectoral girdle (Yan et al., 2005) (Figs. 7G, H). In *sox9a*;*sox9b* double mutants, both the scapulocoracoid and the extension of the fin bud are missing or severely reduced (Yan et al., 2005). We conclude that the reduced expression of *klf2b* and AI722369 in mutant embryos arises from fin bud hypogenesis downstream of Sox9 and so these genes may not be direct targets of Sox9.

Discussion

Because *Sox9* encodes a transcription factor that functions by regulating the transcription of down-stream target genes, the identification of targets is an appropriate way to understand the various roles of *Sox9* in development. Although several

downstream targets of Sox9 are known, many others remain to be identified. A genome-wide screening for the identification of targets provides a fruitful method for recognizing previously unknown targets, and hence for identifying unrecognized functions of Sox9. To implement this strategy, we compared the gene expression profile of wild-type and *sox9*-deficient mutant embryos using a microarray representing 16,228 gene clusters. Because Sox9 generally acts as a transcriptional activator (Südbeck et al., 1996), we focused on genes whose expression was down-regulated in mutant embryos. We identified several types of candidate target genes, including genes expressed in cartilage, retina, and pectoral fin bud. A comparison of *sox9* expression patterns in the retina to the phenotypes of *sox9* mutants and the expression patterns of candidate targets in wild-type and *sox9* mutant embryos uncovered a previously unrecognized requirement for *Sox9* in development of the retina.

Down-stream targets of Sox9

Microarray is a powerful tool to identify genes whose expression is different among experimental groups (Schena et al., 1995; Hoheisel, 2006). Expression differences between groups could arise due either to direct, primary effects, or to subsequent, secondary effects. A direct target of Sox9 should be co-expressed with *sox9* in space. A secondary target would be turned on in response to a primary target in a regulatory cascade, and so might be first expressed well after *sox9* or be expressed in cells that do not express *sox9*. To confirm microarray results, we compared the expression patterns of candidate genes and *sox9* genes among wild-type, *sox9a*, *sox9b* and *sox9a*;*sox9b* double mutant embryos by whole mount in situ hybridization. Our microarray experiments identified potential Sox9 target genes in two major developmental processes,

skeletogenesis and retinogenesis, as well as several genes likely to represent secondary effects.

Targets involved in skeletogenesis

Microarray analysis showed that the collagen genes *col2a1a* and *col11a2* were down-regulated in mutants. Because the mouse and human orthologs of *col2a1a* and *col11a2* are well characterized direct targets of Sox9 (Bell et al., 1997; Lefebvre et al., 1997; Ng et al., 1997; Bridgewater et al., 1998; Bi et al., 1999; Sekiya et al., 2000), these genes validate the method. Indeed, it has already been shown in zebrafish embryos that the expression of *col2a1a* is down-regulated in *sox9a*, *sox9b* and double mutant embryos (Yan et al., 2005).

This study presents the first description of *col11a2* expression in zebrafish. At 2dpf, *col11a2* was expressed in neurocranial cartilages, pharyngeal cartilages, otic vesicle and pectoral fin bud. This expression pattern is consistent with mouse *Col11a2*, which is expressed in cartilaginous tissues, such as chondrocranium, hyoid, Meckel's, and limb cartilages (Sugimoto et al., 1998). Our results showed that the expression of zebrafish *col11a2* was affected differently in *sox9a*, *sox9b*, and *sox9a*;*sox9b* double mutant embryos. Thus, we conclude that *sox9a* and *sox9b* both regulate *col11a2* expression in tissues in which each *sox9* gene is expressed. After the *sox9* gene duplication event, regulatory elements controlling the expression of *sox9* co-orthologs apparently partitioned in a tissue-specific fashion, but the Sox9 proteins encoded by *sox9a* and *sox9b* retained the ability to regulate *col11a2*, presumably by binding to regulatory sites for the *col11a2* gene. Other known or potential Sox9 targets, like *aggrecan*, *hyaluronan* and *proteoglycan link protein 1b* (*hapln1b*), *runx2a* and *runx2b* were not spotted on the

array.

Targets involved in retinogenesis

Microarray analysis detected down-regulation of genes expressed in the retina, including *calb2a*, *calb2b*, *crx*, *neurod*, *rs1*, *sox4a* and *vsx1*; whole mount in situ hybridization experiments confirmed reduced expression of these genes in mutant retinas. Our analysis revealed the detailed expression pattern of *calb2a*, *calb2b*, *rs1* and *sox4a* in developing retina, which had not previously been characterized in detail, although some data had been deposited in the high-throughput expression database in ZFIN (Rauch et al., 2003; Thisse and Thisse, 2004).

In situ hybridization results for *calb2b* were similar to the previously reported antibody staining for the Calretinin, which labels ganglion, amacrine and bipolar cells (Bernardos et al., 2005). Although *calretinin* is an alternative name for *calb2a* in zebrafish, the popularly used anti-Calretinin polyclonal antibody (AB5054) seems likely to recognize also Calb2b in zebrafish. Mouse Calretinin (Calb2) is a marker for ganglion and amacrine cells in retinas, and the same cell types express *calb2b* in zebrafish. Expression results suggest evolutionary models: the ancestral expression domain of *Calb2* was in ganglion and amacrine cells of the retina, and that this pattern was preserved by *Calb2* in mouse and by *calb2b* in zebrafish, while zebrafish *calb2a* might have acquired an expanded expression domain in the inner nuclear layer of the retina. Another possible model would be that the ancestral *Calb2* was expressed in a pattern that equals the sum of the zebrafish *calb2a* and *calb2b* domains, that these domains partitioned between two zebrafish *calb2* genes, and that the mouse lineage preserved only the ganglion and amacrine expression domains.

Human *RS1* is associated with X-linked retinoschisis, a common form of macular degeneration (Sauer et al., 1997). *Rs1* encodes a protein containing a single discoidin domain shared with receptors for collagen (Vogel, 1999; Vogel et al., 2006). Although it has not been shown that Rs1 protein interacts with collagen, Rs1 is interesting as a candidate target for Sox9 because the collagens *Col2a*, *Col9a1* and *Col11a2* are known as Sox9 targets (Bell et al., 1997; Lefebvre et al., 1997; Ng et al., 1997; Bridgewater et al., 1998; Bi et al., 1999; Sekiya et al., 2000; Zhang et al., 2003; Genzer and Bridgewater, 2007). Mouse *Rs1* is down-regulated in the retina of *Crx* knockout mice (Livesey et al., 2000; Blackshaw et al., 2001). The down-regulation of zebrafish *rs1* in *sox9* mutant retinas could thus either be a secondary effect due to the primary down-regulation of *crx*, or it might directly result from regulation by Sox9.

In zebrafish, *crx* is expressed in proliferating retinal progenitors and its expression was diminished in *sox9b* mutants. Our preliminary comparative genomic analyses identified a conserved non-coding element (CNE) near the *crx* gene that has Sox9 consensus binding sites (data not shown). Although we haven't confirmed its enhancer activity on *crx* gene expression, it is possible that Sox9 binds to these cis-elements and directly activates *crx* gene expression.

Our result showed that expression of *vsx1* became down-regulated in *sox9b* mutants, and it has been published that expression of *vsx1* also down-regulates in zebrafish *crx* morpholino injected embryos (Shen and Raymond, 2004). Thus, the reduction in *vsx1* expression in *sox9b* mutant embryos could result from either a direct effect of *sox9b* loss or a secondary effect caused by the reduced *crx* activity that is associated with loss of *sox9b*. Both *crx* and *vsx1* have predicted Sox9 binding sites (data

not shown), as expected by the hypothesis that they are directly regulated by Sox9.

A novel function for Sox9 in retinal development

The retina of zebrafish, human, and other vertebrates consists of seven major cell types: six types of neurons and a glial cell type, all of which differentiate from the neural epithelium (Malicki, 2000; Stenkamp, 2007). In zebrafish, retinal neurogenesis proceeds sequentially; ganglion cells are generated from 24 to 36 hpf, cells of the inner nuclear layer arise from 36 to 48 hpf, and outer nuclear layer cells are born from 48 to 60 hpf (Stenkamp, 2007). Differentiation initiates in the ventral retina and then spreads dorsally around over time (Fadool and Dowling, 2008; Stenkamp, 2007). One of the zebrafish *sox9* genes, *sox9b*, begins to be expressed in the eye by 24 hpf (Chiang et al., 2001), and at 2 dpf, localizes to the ciliary marginal zone of the retina, where retinal progenitors proliferate. In the retina of *sox9b* mutants, expression of candidate genes was severely reduced at 2 dpf in the outer nuclear layer, inner nuclear layer and ganglion nuclear layer, suggested that *sox9b* is required for differentiation of most of the retina. Although the expression of candidate genes became milder at 3 dpf, gene expression close to the ciliary marginal zone was still reduced, and Müller glial cells and photoreceptor cells were diminished in number and were disorganized in location at 5 dpf. Considering the expression pattern of *sox9b*, the expression of candidate genes in *sox9b* mutants, and the phenotype observed in the retina of *sox9b* mutants, we conclude that *sox9b* is essential for retinal development. The requirement of *sox9b* for retinal development is most prominent at 2 dpf, and it becomes less essential at 3 dpf but still required for organization and proliferation of Müller glial cells and photoreceptor cells; thus, these results uncover a novel function for *Sox9*.

Evolution of functions of zebrafish Sox9 co-orthologs

After gene duplication, preserved duplicates often experience reciprocal loss of gene subfunctions (Force et al., 1999; Postlethwait et al., 2004), for example, the zebrafish *Sox9* co-orthologs, *sox9a* and *sox9b*, appear to have partitioned the ancestral craniofacial skeleton expression domain between the two genes. In other domains, both duplicates preserve ancestral functions, for example, both *sox9a* and *sox9b* are expressed in the otic vesicle. Because *Col11a2* is a direct target of Sox9 (Bridgewater et al., 1998; Liu et al., 2000), and also because *col11a2* was isolated as a candidate *sox9* target in this study, the expression of *col11a2* would be a good marker for visualizing functions of *sox9*, and for tracing the partitioning of subfunctions between *sox9a* and *sox9b*.

In the neurocranium, the ethmoid plate expresses *sox9a* but not *sox9b*, and the ethmoid plate is deleted in *sox9a* mutants but not in *sox9b* mutants; thus, our results show that *col11a2* expression in the ethmoid plate depends on *sox9a* but not *sox9b*. On the other hand, wild-type pharyngeal arches express both *sox9a* and *sox9b*, with *sox9a* in the mesenchyme and *sox9b* in the epithelium (Yan et al., 2005), but neither *sox9a* nor *sox9b* single mutant embryos express *col11a2* in pharyngeal arches. This result shows that the activity of both *Sox9* co-orthologs is required in the pharyngeal arches to regulate *col11a2*. Both *sox9a* and *sox9b* are expressed in the otic vesicles, but in different, though overlapping domains (Chiang et al., 2001; Yan et al., 2005). Likewise, *sox9a* and *sox9b* single mutant embryos both express *col11a2* in the otic vesicle, but in somewhat different patterns, and the function of both genes must be missing to ensure that the entire ear is gone. This result shows that, like the pharyngeal arches, the two *Sox9* co-orthologs likely induce *col11a2* expression in the domains of the otic vesicle in which they are expressed.

We conclude that the regulatory relationship between *sox9* genes and *col11a2* differs among tissues: *sox9a* is sufficient for *col11a2* expression in the ethmoid plate, both *sox9a* and *sox9b* are essential in pharyngeal arches, and each of the duplicates controls *col11a2* expression in different parts of the otic vesicles.

Although *Sox9* function in the eye is not yet well characterized in mouse or human, in mouse, Sox9 protein is expressed in eye (Ihanamäki et al., 2002), and human patients heterozygous for dominant *SOX9* mutant alleles sometime experience reduced eye function (Mansour et al., 2002). These results suggest that the requirement of *Sox9* in retinal development is conserved among vertebrates.

We propose that a requirement for *Sox9* function in retinal development already existed in the last common ancestor of tetrapods and teleost fish, and that after the genome duplication in the teleost fish lineage, retinal function mostly partitioned to *sox9b*, which now acts to control cell numbers and organization during retinal development by activating, directly or indirectly, genes identified by our expression profiling study. Future experiments should be directed towards the identification of specific DNA elements responsible for this newly discovered role of *Sox9*.

CHAPTER III

FUNCTIONAL GENOMICS IDENTIFIES NEURAL STEM CELL SUB-TYPE EXPRESSION PROFILES AND GENES REGULATING NEUROBLAST HOMEOSTASIS

This work was submitted to the journal *Developmenal Biology* in September 2011 and is currently 'under review'. Myself, Kristin Robinson, and Travis Carney performed the microarray-based expression analysis. Travis Carney, Jessica Osterhout, and myself performed the RNAi analysis. Omer Bayraktar performed the immunohistochemistry analysis. Chris Doe was the principle investigators for this work.

Introduction

Drosophila neuroblasts are a powerful model system for understanding the molecular control of stem cell self-renewal versus differentiation. The majority of neuroblasts (type I neuroblasts) repeatedly divide asymmetrically with respect to size and fate to self-renew and produce a smaller daughter cell called a ganglion mother cell (GMC) that divides only once to produce two post-mitotic neurons or glia (reviewed in: Chia et al., 2008; Doe, 2008; Knoblich, 2008). Neuroblast/GMC fate differences are due in part to the asymmetric partitioning of proteins into the GMC during neuroblast cell division. These factors include the transcription factor Prospero (Pros), the Notch inhibitor Numb, and the translational repressor Brain tumor (Brat) (Betschinger et al., 2006; Broadus et al., 1998; Hirata et al., 1995; Knoblich et al., 1995; Lee et al., 2006c; Spana and Doe, 1995). Proper segregation of Pros, Numb, and Brat into the GMC require the scaffolding protein Miranda (Mira) and the WD40-domain protein Lethal giant larvae (Lgl) (Betschinger et al., 2006; Lee et al., 2006c; Ohshiro et al., 2000; Peng et al., 2000).

In the GMC, Pros enters the nucleus and promotes cell cycle exit and differentiation by directly activating differentiation genes and repressing self-renewal and cell cycle regulatory genes (Choksi et al., 2006; Li and Vaessin, 2000). This process allows for a single neuroblast to generate a lineage of many differentiated neurons and glia and for a relatively small number of neuroblasts to generate the thousands of cells found in the central nervous system of the adult fly.

Recently, "type II" neuroblasts were identified in the larval brain which divide asymmetrically to produce small transit-amplifying progenitors (intermediate neural progenitors, INPs). INPs themselves undergo molecularly asymmetric cell divisions to generate 4-6 GMCs, each of which typically generates two post-mitotic neurons or glia (Bello et al., 2008; Boone and Doe, 2008; Bowman et al., 2008; Izergina et al., 2009; Viktorin et al., 2011). Six type II neuroblasts inhabit the dorso-medial region of the lobe and are designated DM1-6, and two occupy more lateral positions (Bayraktar et al., 2010; Izergina et al., 2009). Type II neuroblasts behave in a manner similar to mammalian neural stem cells in that they generate transit-amplifying INPs. Transit-amplifying progenitors are important in the development of the nervous system in mammals as well as in flies because they permit the rapid amplification of neuronal progeny (Bello et al., 2008; Boone and Doe, 2008; Bowman et al., 2008; Merkle and Alvarez-Buylla, 2006; Morrison and Kimble, 2006). Thus, while there are ~95 type I neuroblasts per larval brain lobe and only eight type II neuroblasts, these few lineages produce a considerable fraction – approximately a quarter – of the neurons of the adult brain (Izergina et al., 2009). Type II neuroblasts and INPs have thus become a model for the study of transit-amplifying neural progenitors, and determining how these cells are specified and

maintained may shed light on the function of mammalian neural stem cells.

Type II neuroblasts are known to differ from type I neuroblasts in several ways. First, they generate INPs and thus make much larger lineages than type I neuroblasts (Bello et al., 2008; Boone and Doe, 2008; Bowman et al., 2008). Second, they are the sole contributors of the intrinsic neurons of the adult central complex (Bayraktar et al., 2010; Izergina et al., 2009). Third, they are more susceptible to tumor formation (Boone and Doe, 2008; Bowman et al., 2008; Ouyang et al., 2011; Weng et al., 2010). Numerous genotypes have been identified that cause the production of ectopic larval brain neuroblasts, and several of these specifically affect type II neuroblasts. For example, mutations in *brat* lead to "overgrowth" of just the type II neuroblasts, and mutations in *lgl* affect type II much more strongly than type I neuroblasts. On the other hand, loss of *Aurora-A* (*Aur*) or neuroblast misexpression of membrane-tethered atypical protein kinase C ($aPKC^{CAAX}$) leads to ectopic type I and type II neuroblasts (Bowman et al., 2008). It is likely that the lack of Pros in the new-born INP renders this daughter cell sensitive to the loss of a second growth inhibitor/differentiation factor, making it easier for this cell to revert to a type II neuroblast identity.

In spite of the marked differences between type I and type II neuroblasts in proliferative potential and susceptibility to tumor formation, only two molecular differences are known: type II neuroblasts lack the transcription factors Pros and Asense (Ase), while both are present in all type I neuroblasts (Bello et al., 2008; Boone and Doe, 2008; Bowman et al., 2008). It is currently unknown how many other genes are regulated differentially between type I and type II neuroblasts, and which of them regulates each distinct aspect of type I and type II function. In order to discover such transcriptional

differences, relatively pure populations of each neuroblast sub-type must be isolated from which to extract RNA. Complicating these efforts, the Drosophila central nervous system contains only a small number of neuroblasts which are dispersed throughout a complex population of thousands of neurons and glia, making it difficult to physically separate neuroblast sub-types from each other and from other cell types. Thus, comparing the transcriptional outputs of neuroblast sub-types is technically challenging due to the difficulty of isolating cell type-specific RNA.

In order to enrich for each type of neuroblast, here we make use of published and unpublished mutants in which type I and type II neuroblasts exhibit differential overproliferation phenotypes. We perform microarray-based whole-genome transcriptional profiling to compare each of these different mutant brains to wild-type; thus we are able to probe the transcriptional differences not only between each mutant and wild-type, but also between type I and type II neuroblasts. We identify only a small number of genes exhibiting transcriptional differences between type I and type II neuroblasts, providing a highly specific group of genes to screen for a function in establishing each type of neuroblast. We identify a large group of genes which are likely expressed in neuroblasts but not neurons, and we verify the neuroblast function of a subset of these genes using an RNAi-based targeted loss of function screen. Using this approach we identify 84 genes required to maintain neuroblast numbers in larval brains, all of which have conserved mammalian orthologs.

Methods

Fly stocks

All fly stocks used in this study have been previously described except for *lgl⁻ lgd^{d7},* in which the *lgl* locus has been spontaneously lost (Jason Boone, unpublished data) from the *lgd^{d7}* chromosome (Jaekel and Klein, 2006). Other fly stocks used were: *lgl^{334}* (Rolls et al., 2003); *aurA^{8839}* (Lee et al., 2006a); *lgl^{334};pins^{62}* and *UAS-aPKC^{CAAX}* (Lee et al., 2006b); *brat^{11}* (Lee et al., 2006c); *wor-gal4* (Albertson et al., 2004); *R9D11-Gal4* and *R19H09-Gal4* (Bayraktar et al., 2010); *UAS-mCD8::GFP* and *UAS-Dicer2* (Bloomington Stock Center).

Microarray analysis

Mutant larvae were dissected at 144 hours after larval hatching (ALH); wild-type heterozygous larvae that were at the same developmental stage were dissected at 96 hours ALH as wild-type controls. The only exceptions were the *aPKC^{CAAX}* experiments, in which the experimental larvae were raised at 30°C and exhibited enlarged brains packed with ectopic neuroblasts, and genetically identical control larvae were raised at room temperature, where the ectopic neuroblast phenotype is much weaker. Total RNA was extracted from larval brain lobes using TRIzol extraction methods according to the manufacturer's instructions. First strand synthesis and amplification of Cy3- and Cy5-labeled RNA were accomplished using the Agilent Low Input Quick Amp Labeling Kit. Four experimental replicates were performed for each mutant genotype: two standard replicates (Cy5-labeled mutant RNA and Cy3-labeled wild-type RNA) and two dye-swapped replicates. Exceptions were the *aPKC^{CAAX}* experiments, in which one standard and two swapped replicates were used for clustering, and the *lgl* experiments, in which

three standard and one swapped replicates were used. Hybridization was performed as previously described (Miller et al., 2009), except for each replicate, 825 ng of both mutant and wild-type RNA were mixed and hybridized to Agilent microarrays. The slides were scanned using an Axon GenePix 4000B, and GenePix software was used for feature extraction.

Cluster analysis

We selected genes for cluster analysis if, in at least one of the mutant genotypes, their average transcript levels over all biological replicate experiments deviated from the wild-type RNA sample by greater than two-fold. This criterion resulted in the selection of 2781 genes. To investigate the reproducibility of replicate experiments of the same mutant genotype and the overall transcriptional similarities between experiments, we first performed cluster analysis without averaging individual replicate experiments. This analysis allowed us to determine whether, for each experiment, the clustering analysis grouped replicates of the same genotype together. To identify groups of genes with similar transcript patterns over the different mutant genotypes, we averaged replicates and performed cluster analysis on these averages in order to avoid artificial gene clustering relationships due to technical noise between experimental replicates. After performing cluster analysis, we found that a large group of genes with increased expression in mutant brains clustered with high correlation. Group A was defined by a tree branch that represented a large decrease in correlation, from >0.6 to ~0.38. We similarly found that a long branch in the genes with reduced expression in mutant brains caused a decrease in correlation that passed the .6 cutoff (from >0.7 to ~0.54); thus we considered all the genes clustering below this branch to comprise group C. The remainder

of the genes exhibited variable expression patterns and did not cluster with high correlation, and was defined as group B.

RNAi screen

Knockdowns were performed on group A genes with annotated human orthologs and transgenic RNAi stocks available from the Vienna Drosophila RNAi Center (VDRC). We used these lines to knock down genes cell type-specifically by crossing RNAi line males to *wor-Gal4UAS-Dicer2* virgins. The progeny of these crosses were raised at 30°C and scored for lethality. Each knock-down was performed at least twice to judge the consistency of the phenotype. For those RNAi constructs that caused lethality, we performed crosses again and dissected brains from wandering third instar larvae. We took confocal stacks of these brains (see below) and determined the number of central brain neuroblasts per brain lobe using antibodies against the neuroblast-specific proteins Deadpan (Dpn) and Mira. Optic lobe neuroblasts were excluded from these counts based on their small size, tight clustering, and stereotyped lateral position in the brain lobe.

Fixation, antibody staining, and confocal microscopy

The following antibodies were used for immunohistochemical staining of larval brains: guinea pig anti-Dpn, 1:2000 (J. Skeath); rat anti-Dpn, 1:1-1:50 (Doe lab); rat anti-Elav, 1:50 (Developmental Studies Hybridoma Bank 7E8A10); rabbit anti-Ase, 1:2000 (Brand et al., 1993); rabbit anti-Optix, 1:500 (Kenyon et al., 2005); rabbit anti-Rx, 1:2000 (Davis et al., 2003); chicken anti-GFP, 1:1000 (Aves Laboratories); mouse anti-Pros, 1:1000 (MR1A). Brains were dissected in Schneider's medium, fixed in PBST (phosphate-buffered saline + 0.1% Triton-X100) with 4% formaldehyde for 20 minutes, rinsed for 30 minutes in PBST, blocked for 30 minutes using PBSBT (PBST + 1% bovine

serum albumin) or PBST + 5% normal goat serum. Brains were incubated in primary antibody overnight at 4°C with rocking, and then rinsed in PBST + block for 1 hour. Brains were incubated with secondary antibodies (1:500; Molecular Probes or Jackson Immunoresearch) for 2 hours at room temperature with rocking, and then rinsed for 1 hour with PBST and stored in Vectashield until microscopy could be performed. Microscopy images were taken using a Bio-Rad Radiance or Zeiss700 confocal microscope.

Results

Transcriptional profiling of larval brains containing ectopic type I or type II neuroblasts

We analyzed six different genotypes that generate ectopic neuroblasts in the third instar larval brain (Table 2). The *brat* and *lgl* single mutants produce primarily ectopic type II neuroblasts, whereas *aur* mutation or misexpression of membrane-tethered aPKC is reported to generate ectopic type I and type II neuroblasts (Bowman et al., 2008). We also analyzed *lgl lgd* and *lgl pins* double mutants, both of which produce large numbers of ectopic neuroblasts of unknown type (Lee et al., 2006a; Lee et al., 2006b; Wang et al., 2006; Jason Boone, unpublished data). We stained these brains for the pan-neuroblast marker Dpn and neuronal marker Elav to determine the total number of ectopic neuroblasts and remaining number of neurons, showing that there is a graded increase in the number of neuroblasts per brain from *lgl* (the fewest ectopic neuroblasts) to *lgl pins* (almost entirely neuroblasts; Figs. 8A and 8B). As type I and type II neuroblasts can be distinguished by the presence of Ase only in type I neuroblasts (Figs. 8C and 8D), we

also stained the brains for Dpn and Ase to determine the proportion of ectopic type I/type II neuroblasts. We confirmed that the *brat* single mutant generates primarily ectopic type II Ase^- neuroblasts while *aur* and $aPKC^{CAAX}$ brains contain more type I Ase^+ neuroblasts, although there is also an increase in type II neuroblasts (Fig. 8E). Moreover, we found that the *lgl lgd* double mutant is strongly enriched for type II neuroblasts, and *lgl pins* brains contain both neuroblast types with an enrichment of type I neuroblasts. We noted that in *lgl pins* brains, distinct regions of type I and type II neuroblast overproliferation are discernible based on the lack of Ase and Pros in ectopic cells derived from type II neuroblasts (Fig. 8E and inset). We conclude that the six genotypes used here exhibit a range of type I/type II differential overproliferation phenotypes, with *brat, lgl,* and *lgl lgd* brains representing enriched pools of type II neuroblasts, and with *aur*, $aPKC^{CAAX}$, and *lgl pins* being more enriched for type I neuroblasts.

Next, we used transcriptional profiling of the larval brain lobes from each of these six genotypes to identify (a) genes differentially regulated in type I vs. type II neuroblasts, that may function in establishing the striking differences between these two types of progenitors, and (b) genes expressed in all neuroblasts, that may function to regulate self-renewal or asymmetric cell division. For each experiment, we isolated RNA from mutant and wild-type brain lobes, amplified and fluorescently labeled each RNA sample, and hybridized them directly against each other to microarrays representing the entire complement of protein-coding Drosophila genes with at least two-fold redundancy (Fig. 8F). We used cluster analysis to group genes according to transcriptional pattern similarities in the different experiments. Genes exhibiting no change between wild-type and mutant were not included in the cluster analysis (see Methods). Biological replicates

and dye-swap experiments cluster much more closely to one another than to replicates for any other mutant (data not shown); this demonstrates that our data are highly reproducible and that each mutant exhibits a distinct transcriptional profile.

We sorted genes into three groups based on their transcript pattern in the six genotypes (Fig. 9A). Group A contains 1045 genes with elevated expression in the mutant genotypes (Fig. 9B); these genes are likely to be expressed in neuroblasts and are good candidates for regulating neuroblast function (see below). Group B contains 467 genes that have variable expression between mutants. Group C contains 1269 genes with decreased transcript levels in each mutant (Figs. 9A and 9B); thus the genes in this large group are good candidates for genes expressed in neurons or glia but not in neuroblasts. Interestingly, we did not see clustering of the genotypes that generate ectopic type II neuroblasts (*lgl, lgl lgd*, and *brat*) – note that the dendrogram at the top of Fig. 9A shows that each of these genotypes has a more closely related genotype that generates ectopic type I neuroblasts – suggesting type I and type II neuroblasts are much more similar transcriptionally than different.

Identification of genes transcribed preferentially in type I or type II neuroblasts

To identify genes expressed differentially between type I and type II neuroblasts, we looked for genes clustered with *pros* and *ase*, the only two genes known to be differentially expressed in type II neuroblasts. We found that *pros* and *ase* reside together in a small sub-cluster of only 11 genes within group B (Fig. 10A). This sub-cluster as a whole exhibits reduced expression in *brat*, *lgl,* and *lgl lgd* mutants and enrichment in *aur, aPKC*^{CAAX}, and *lgl pins;* remarkably, no other sub-cluster exhibits such a pattern. This suggests that the other nine genes in the cluster may also be specifically expressed in type

I neuroblasts, like *pros* and *ase,* and that these are potentially the only genes that exhibit this unique pattern.

To test whether other genes in the small *pros/ase* cluster are also expressed in type I neuroblasts but not type II neuroblasts, we obtained an antibody to a candidate from this cluster, Retinal homeobox (Rx), a homeodomain-containing transcription factor (Davis et al., 2003; Eggert et al., 1998). We found that Rx is completely absent from type II neuroblasts, similar to Pros and Ase; Rx is detected in several type I neuroblasts as well as in a subset of differentiated type II progeny (Figs. 10B and 10C). Consistent with this expression pattern, we found that *brat* mutants, which overproduce type II neuroblasts, show a loss of Rx staining (Fig. 10D). In contrast, *lgl pins* mutants, which have ectopic type I neuroblasts, show territories of strong Rx expression which is confined to $Pros^{+}$ (likely type I-originating) cells (Fig. 10E). The fact that only a small patch of *lgl pins* mutant brain tissue is Rx^{+} is probably because Rx is normally expressed in a subset of type I neuroblasts. We conclude that Rx, like Pros and Ase, is expressed in type I but not type II neuroblasts. Thus, most or all of the 11 genes in the *pros/ase* sub-cluster may be expressed in type I but not type II neuroblasts.

We next wanted to find genes expressed in type II neuroblasts but not type I neuroblasts, as there are currently no known markers specifically expressed in type II neuroblasts. We reasoned that transcripts expressed in type II neuroblasts should be enriched in genotypes that overproduce type II neuroblasts: *brat, lgl* and *lgl lgd*. We found one small cluster enriched in two of the three mutants (*brat* and *lgl lgd*) (Fig. 11A). This cluster contains just six genes, all encoding transcription factors. To verify the expression pattern of this gene cluster, we examined the expression of one gene product,

Optix. Optix is a conserved homeodomain-containing transcription factor required for eye development (Kenyon et al., 2005; Seimiya and Gehring, 2000; Toy et al., 1998). Consistent with our microarray data, we found that most of the Optix expression in the brain is indeed restricted to type II lineages; four of the six dorso-medial type II neuroblasts (DM1, 2, 3, and 6) express Optix, as do most of the INPs, GMCs, and neurons in these lineages (Figs. 11B and 11C). The other two dorso-medial type II lineages (DM4 and 5) exhibit some expression of Optix in a subset of neuronal progeny, but it is absent from the neuroblasts and INPs in these lineages (Figs. 11B and 11C and not shown). In addition, a single dorsal type I neuroblast expresses Optix (Fig. 11C). Inspection of mutant brains further confirmed the type II-biased expression of Optix, in that *brat* mutant brains exhibit a marked increase in $Optix^{+}$ neuroblasts (Fig. 11D), and in *lgl pins*, the increase in Optix is almost exclusively in a $Pros^{-}$ (type II-originating) region of the brain (Fig. 11E). Our results indicate that our clustering relationships can be used to predict type I/type II expression bias with good accuracy. We conclude that Optix is primarily expressed in type II but not type I neuroblasts, and that Optix and the other five genes in this cluster are excellent candidates for regulators of type II neuroblast identity.

Identification of genes predicted to be expressed in neuroblasts but not neurons

To determine whether the ectopic neuroblasts in the mutant brains express wild-type neuroblast genes, we tested whether genes known to be expressed primarily or exclusively in neuroblasts are found in group A; indeed all such positive control genes (with the exception of the type II-negative *pros* and *ase* genes) are represented in group A, including *worniu* (*wor*)*, deadpan,* and *CyclinE* (Table 3). Conversely, neuronal and glial genes are excluded from group A and found in group C [e.g. *elav, glial cells missing*

(*gcm*), and *reversed polarity* (*repo*)] (Table 3). In addition, there is a good correlation between the number of neuroblasts in each mutant brain and the level of enrichment shown for group A genes (Figs. 8B and 8B). Thus, the genes in group A are likely to be expressed in both type I and type II neuroblasts, but not in neurons or glia. Conversely, genes in group C are likely to be expressed in differentiated neurons or glia, but not in neuroblasts. We conclude that ectopic neuroblasts are similar transcriptionally to wild-type neuroblasts, and thus the mutant genotypes represent an enriched source of neuroblast-expressed mRNA.

We next determined the gene ontology (GO) terms that represent the neuroblast-enriched group A genes and the neuron/glia-enriched group C genes. We found that group A genes are strongly enriched for several GO terms, including cell cycle and ribosome biosynthesis – processes expected in neuroblasts that must repeatedly divide and grow (Figs. 12A and 12B). For example, a small sub-cluster in group A exhibits a very significant enrichment for genes involved in DNA replication ($p<10^{-15}$); this process is not significantly represented in any other cluster ($p>.01$ for all other group A genes combined; Fig. 12A). Conversely, we found group C to be significantly enriched for the GO terms morphogenesis, signal transduction, and differentiation – all expected for post-mitotic neurons and glia (Fig. 12B). In addition, neuropeptide signaling and cell morphogenesis are both significantly enriched ($p<10^{-20}$ and $p<10^{-21}$, respectively) in distinct sub-clusters (Fig. 12C). We conclude that group A is enriched for genes that are expressed in neuroblasts but not differentiated neurons and glia, and group C is primarily composed of genes expressed in post-mitotic neurons and glia.

Functional analysis of genes predicted to be expressed in neuroblasts but not neurons

To determine if group A genes are required for neuroblast survival, proliferation, or self-renewal, we performed RNAi knock-down experiments. We selected genes for which transgenic RNAi stocks were available from the Vienna Drosophila RNAi Center, and we further restricted our analysis to those genes with human orthologs in order to enhance the relevance of our study to issues of human stem cell function (Fig. 13). This resulted in our analyzing 691 RNAi lines representing 595 genes. We reasoned that loss of function of genes with critical functions in neuroblasts would cause defective central nervous system development and eventual lethality, as seen in other genotypes which affect neuroblast function. Thus we screened for lethality in the progeny of males from each RNAi line crossed to *wor-Gal4 UAS-Dicer2* flies at 30°C [*wor-Gal4* drives expression in neuroblasts (Albertson et al., 2004), while *Dicer2* improves RNAi efficacy (Dietzl et al., 2007)]. Of the 691 RNAi lines tested, 195 (28%) cause lethality or semi-lethality (Fig. 13A). We found that of the genes for which we tested multiple RNAi lines, 84% exhibit the same lethality phenotype for both lines. Few RNAi lines cause embryonic lethality at 30°C, and in these cases larval stages were obtained by setting up crosses at 18°C and shifting larvae to 30°C after embryogenesis. The lack of a lethal phenotype in 72% of the lines may be due to either inefficient RNAi knock-down of gene expression or the non-essential function of the gene in larval neuroblasts. Hence we restricted our subsequent analysis to the 28% of lines with a lethal or semi-lethal phenotype.

We tested each of the lethal or semi-lethal genes for a change in neuroblast

number, reasoning that genes expressed in neuroblasts but not neurons may play a role in neuroblast survival, quiescence, identity, asymmetric division, or self-renewal. We performed the same RNAi experiments as above and determined the number of Dpn^+ $Mira^+$ central brain neuroblasts (optic lobe neuroblasts were not assayed). We found that nearly one half of lethal genes (86) cause a significant change in central brain neuroblast numbers (Fig. 13B, Fig. 13C, and data not shown). A majority of these changes are decreases in neuroblast number, as expected based on the predicted expression of these genes in neuroblasts. Two genes known to regulate neuroblast numbers were detected in the screen, *mira* and *aurora borealis*, thereby validating this approach. Importantly, all of these genes have clear mammalian orthologs. We conclude that our RNAi-based screening method has yielded a list of 84 new candidates for regulating neuroblast self-renewal.

Discussion

It has previously been shown that co-clustering of genes in expression profiling data is likely to reflect physical or genetic interactions (Ge et al., 2001; Jansen et al., 2002) and participation in the same pathways (van Noort et al., 2003). Our results are consistent with these conclusions. For example, we identified a small group of 11 genes containing the only two genes known to be expressed in type I but not type II neuroblasts, and showed that a third gene has a similar pattern of expression – thus all genes in this cluster are likely to be expressed in type I but not type II neuroblasts. Furthermore, the strong enrichment of GO terms in small sub-clusters within both group A and group C (Fig. 12) indicates that genes within these clusters are likely to share similar functions or

processes.

Differences between type I/type II neuroblasts are caused by a small number of genes

At the outset of this study, we expected to find a large group of genes that were differentially expressed in type II versus type I neuroblasts, because these neuroblasts have such strikingly different cell lineages. However, we were only able to identify a few gene clusters that were differentially regulated in such a type I/type II consistent manner – the 11 genes in the *pros/ase* cluster depleted in type II neuroblasts and the six genes enriched in type II neuroblasts (Figs. 10A and 11A). This suggests that the small number of genes that we identified may play a disproportionately large role in generating differences between type I and type II neuroblasts. Might *pros* and *ase* be the only genes regulating type I/type II differences? Both Ase and Pros can promote cell cycle exit (Choksi et al., 2006; Dominguez and Campuzano, 1993; Li and Vaessin, 2000; Southall and Brand, 2009; Wallace et al., 2000), which may result in the Ase^{+} $Pros^{+}$ type I progeny taking a GMC identity and undergoing just one terminal division and the Ase^{-} $Pros^{-}$ type II progeny taking an INP identity and continuing to proliferate. Indeed, the misexpression of either Ase or low levels of Pros in type II neuroblasts is sufficient to cause the loss of INPs and/or their premature cell cycle exit, thereby decreasing lineage size toward the size of type I neuroblasts (Bayraktar et al., 2010; Bowman et al., 2008). However, it is unclear what is required to fully transform these cells into type I neuroblasts; addressing this question will require additional molecular markers (some provided by our work here) and tracing the axon projections of the progeny of these "transformed" neuroblasts (e.g. do they now fail to make intrinsic neurons of the adult central complex?). The fact that

mutants in *ase* and *pros* do not transform type I neuroblasts into type II neuroblasts (Bowman et al., 2008; Weng et al., 2010) indicates that other genes, perhaps some in the *pros/ase* cluster described here, are also important for specification of type I neuroblast identity.

Group A: candidate neuroblast-specific genes and neuroblast homeostasis regulators

We found that the neuroblasts in each mutant have remarkably similar expression profiles, as shown by the extensive list of similarly expressed genes in group A and by the list of genes with depleted expression in mutant brains, represented by group C. We believe that these categories provide lists of genes that are representative of those expressed in neuroblasts and neurons, respectively, based on all known neuroblast-specific genes showing up in group A and all known neuron- or glial-specific genes being excluded from group A.

Our RNAi-based screen helped to substantiate this claim, in that a substantial percentage of group A genes caused lethality when subjected to neuroblast-specific knock-down. We do not believe that off-target effects lent a significant amount of error to these lethality data for two reasons: (1) a similar percentage of lethal and non-lethal RNAi lines (about 30%) had more than one non-specific target, which indicates that the observed lethality was due to specific target knock-down; and (2) a majority of genes (about 85%) caused the same lethality phenotype when targeted with multiple independent RNAi lines. Interestingly, we found that many RNAi lines caused lethality with no concomitant change in neuroblast numbers (Fig. 13). We believe this to be due to neuroblast defects which disrupt normal brain development without causing neuroblast loss, per se. For instance, neuroblast failure to make the proper number or type of

progeny might be expected to cause such a phenotype. It will be interesting to investigate the specific effects these essential genes have on neuroblast function. We note that all of the putative regulators of neuroblast homeostasis identified here have mammalian orthologs; these genes are excellent candidates for regulating self-renewal of mammalian neural stem cells.

Group B: expression in subsets of neuroblasts or neurons?

Group B genes apparently are not expressed in all neuroblasts like the group A genes, nor in all neurons or glia like group C genes. However, group B genes are more likely to be expressed in subsets of neurons, not neuroblasts, because group B genes as a whole have an over-representation of GO terms more similar to group C than to group A (Fig. 12B). Why then are group B genes excluded from group C, the neuron cluster? One possible explanation is that different neuroblast lineages are affected in each mutant, and thus different subsets of neurons are missing in each mutant. If different neuroblast lineages express different genes (which seems likely), then each mutant would be missing a unique subset of neural differentiation genes, leading to the cluster being excluded from group C. This model raises the intriguing possibility that group B sub-clusters may represent lineage-specific genes.

It is also possible that the mutant genotypes themselves may cause unique transcriptional differences, leading to a cluster of genes in group B. For example, several small sub-clusters in group B are expressed differently only in $aPKC^{CAAX}$ brains (data not shown). These transcriptional differences are not correlated with the number of type I or type II neuroblasts. Instead, these genes appear to be differentially expressed in response to elevated aPKC. Drosophila aPKC has been best studied as a component of the apical

complex in mitotic neuroblasts, and its capacity for causing ectopic self-renewal has been shown to be reliant on both its catalytic activity and its membrane localization (Atwood and Prehoda, 2009; Lee et al., 2006b). However, aPKC has been ascribed a role in neuroblast proliferation as well as in polarity (Chabu and Doe, 2008; Rolls et al., 2003), and a vertebrate homolog, PKC-ζ, was shown to possess a nuclear role in both proliferation of neural progenitors and neuronal cell fate specification (Sabherwal et al., 2009). These observations are consistent with a role of aPKC in causing transcriptional differences.

Conclusions

Our findings highlight the importance of expression profiling of multiple genotypes. This method allowed us to get a more reliable picture of the group A genes expressed in neuroblasts, because genes with lineage-specific or genetic background-specific changes in expression appeared to be focused into group B, where they do not interfere with the clustering of groups A and C. In addition, we identified two small sub-clusters of genes in group B that are excellent candidates for being preferentially expressed in type I or type II neuroblasts, for which there have been few examples to date. Finally, we conclude that group A genes are likely to be expressed in neuroblasts, and our functional studies have identified 84 genes that are conserved in mammals and required for regulating neuroblast numbers in Drosophila. Future phenotypic analysis in Drosophila will determine whether these genes regulate neuroblast survival, quiescence, asymmetric cell division, and/or self-renewal. Future studies on the expression and function of orthologous genes in mouse neural progenitors and human stem cells (IP or neural) will reveal whether they have conserved roles from flies to mammals.

CHAPTER IV

MOSAIC ANALYSIS WITH DOUBLE MARKERS REVEALS TUMOR CELL OF ORIGIN IN GLIOMA

This work was published in volume 146 of the journal *Cell* in July 2011. Chong Liu performed the immunohistochemistry analysis. Chong Liu, myself, and Roel Verhaak performed the microarray-based expression analysis. Chong Liu and Johnathan Sage performed the quantitative PCR analysis. Simon Hippenmeyer contributed MADM mouse strains. Hui Zong, Liqun Luo, Akiko Nishiyama, Roderick Bronson, Oded Foreman, and Hannes Vogel were the principle investigators for this work.

Introduction

Cancer is a disease of genetic mosaicism because cancerous cells harbor genetic mutations that are absent in normal cells within the same individual. In familial cancer patients, even though initial mutations exist in every cell, in most cases, only specific cell types can progress into malignancy. Those cell types are called cancer cell of origin. Such a cell type-specific susceptibility implies the existence of a permissive or even synergizing signaling context in the cell of origin for particular genetic mutations to cause cancer formation. Therefore, identification of the cancer cell of origin would provide critical insights for understanding tumorigenic mechanisms and for designing rational therapeutic strategies.

Despite such importance, identification of the cell of origin for most cancers has been a daunting task (Visvader, 2011). The reliability of revealing cell identity solely based on molecular and cellular analyses of late stage tumors is often confounded not only by infiltrated bystander cells, but also by the acquired plasticity often found in

terminal cancerous cells. To circumvent these issues, genetically engineered mouse models have been widely used to determine the tumorigenic potential of a specific cell type by initiating mutations with a cell type-specific Cre transgene. However, it is critical to note that cells initially acquiring mutations (cell of mutation) may not be the cell of origin. When mutations are introduced in stem/progenitor cells, it is extremely difficult to distinguish whether initial mutant cells directly transform or whether they merely pass on mutations to more restricted progeny that then transform. In the latter scenario, the mutated stem/progenitor cell is simply the cell of mutation, and the transforming progeny is the actual cell of origin (Visvader, 2011).

The cell of origin for malignant glioma, a type of deadly brain cancer, remains controversial. Successful isolation of tumor cells with stem cell features (known as cancer stem cells) from human gliomas (Singh et al., 2004) implies neural stem cells (NSCs) as the cell of origin. However, such NSC-like features of malignant glioma cells could be acquired during transformation rather than reflect the nature of the original cell type (Visvader, 2011). Further evidence supporting the NSC origin of glioma was obtained from mouse genetic studies. For example, the inactivation of tumor suppressor genes (TSGs) *p53* and *neurofibromatosis 1* (*NF1*) or the expression of a mutant form of *p53* in NSCs consistently led to glioma formation in mouse models, and the physical locations of tumors appeared to associate with the subventricular zone (SVZ), where adult NSCs reside (Alcantara Llaguno et al., 2009; Wang et al., 2009; Zhu et al., 2005). However, other studies suggest that NSC-derived progeny such as astrocytes or oligodendrocyte precursor cells (OPCs) might directly transform (Bachoo et al., 2002; Lindberg et al., 2009; Persson et al., 2010). This unresolved controversy partially stems

from the distinct oncogenic mutations used in these models that make the direct comparison difficult and, more importantly, from the lack of high-resolution analyses of cellular aberrations during the transforming process.

Mosaic analysis with double markers (MADM), a mouse genetic mosaic system (Zong et al., 2005), could, in principle, be used to analyze aberrations in individual cell lineages prior to the final transformation and should thus be suitable for identifying cancer cell of origin. Via Cre/loxP-mediated mitotic interchromosomal recombination, MADM generates a small number of homozygous mutant cells, thus mimicking the sporadic loss of heterozygosity (LOH) of TSGs in human cancers (Knudson, 1971). MADM also permanently labels these mutant cells with green fluorescent protein (GFP) and their sibling wild-type (WT) cells with red fluorescent protein (RFP) within an otherwise unlabeled heterozygous mouse (Fig. 14A). The single-cell resolution benefited from the sparse labeling (0.1%–1% or much lower) (Zong et al., 2005) enables one to track mutant cells throughout the entire process of tumorigenesis. The sibling red WT cells serve as internal controls for green mutant cells, thereby greatly facilitating detailed analyses of cellular aberrations of all lineages in their native environment. In summary, MADM can provide features that are indispensable for a robust analytical paradigm to identify a cell of origin.

Here, we report the application of MADM to glioma modeling. After initiating *p53*/*NF1* mutations sporadically in neural stem cells (NSCs), we analyzed mutant NSCs and all of their progeny at premalignant stages. We found dramatic overexpansion and aberrant growth specifically in OPCs, but not in NSCs or other lineages. Upon tumor formation, marker staining and transcriptome analysis confirmed the OPC nature of

tumor cells. Finally, introducing the same mutations into OPCs consistently led to gliomagenesis. Our findings reveal OPCs as the cell of origin for glioma even when initial *p53*/*NF1* mutations occur in NSCs, thus resolving the current controversy by distinguishing cancer cell of mutation from cell of origin. Importantly, although our studies focused on glioma, the analytical paradigm with MADM that is developed here could be applied to identify cellular origins for many other cancers.

Methods

Mouse lines

Mouse lines used for *NSC-Cre* glioma model: a pair of MADM mice (TG11, GT11) (Hippenmeyer et al., 2010), *hGFAP-Cre* (Zhuo et al., 2001) or *Nestin-Cre* (Petersen et al., 2002), *p53KO* (Jacks et al., 1994) and *NF1flox* (Zhu et al., 2001). Mouse lines used for *NG2-Cre* glioma model: a pair of MADM mice (TG11ML, GT11ML, unpublished lines), *NG2-Cre* (Zhu et al., 2008), *p53KO* (Jacks et al., 1994) and *NF1flox* (Zhu et al., 2001). TG11ML and GT11ML mouse lines have three mutually exclusive loxP variants in MADM cassettes, instead of single loxP sites in TG11 and GT11 mouse lines. This design was used to increase the efficiency of inter-chromosomal recombination induced by *Cre* transgenes.

BrdU labeling

BrdU was administered by i.p. injection (50mg/kg body weight) for short term labeling (1.5hours) or by drinking water (1mg/ml) for long term labeling (7 days). Mice were sacrificed right after the treatment and brain tissue was collected for the further analysis.

Pathology

Tissues used for pathological analysis were collected from O.C.T.-embedded brains. The tissue was thawed and the O.C.T. was removed by washing with cold PBS. The tissues were then preserved in 4% PFA until being processed for paraffin embedding and H&E staining.

Immunohistochemistry

Most immunohistochemistry in our work involved MADM staining (for GFP and RFP), which occupies the commonly used green (488nm laser) and red (543 nm laser) channels in confocal imaging. Considering this, marker staining other than MADM was applied in either the Far-red channel (635 nm laser) or ultraviolet (UV) channel (405 nm laser), or both. If only a single marker was stained along with MADM (which we named three-channel staining, see below), we usually stained this marker with the appropriate secondary antibody to visualize it in the Far-red channel (such as Alexa Fluor 647 or Cy5 conjugated secondary antibodies). If more than one marker besides GFP and RFP was used (we named it as four-channel staining), the fourth would be the UV-channel which was visualized by using appropriate biotin-conjugated secondary antibody to recognize the primary antibody and then Alexa Fluor 405 conjugated streptavidin to develop the fluorescent signal.

Occasionally, we performed staining for three markers along with GFP and RFP (such as those in Figs. 16, we named it as “4+1” channel staining). In this way, we borrowed the red channel from RFP for visualizing the fifth marker. Although this fifth marker staining cannot be visualized in either red WT or double-color cells, it helps us identify marker expression in green mutant cells, especially within tumor region or

among the pre-transforming OPC population where cells are almost exclusively green or colorless. For brain tissue containing grafted tumor cells (secondary tumors) the red channel was saved for staining cell markers other than RFP because no RFP labled WT cells existed in the tumor.

Staining procedures

For three-channel staining with MADM and one more marker, tissue sections were air-dried at room temperature for at least 1 hr after cryosectioning. Tissue sections were then re-hydrated and washed three times (10 min each) with phosphate-buffered saline (PBS) to remove sucrose prior to incubation with blocking/ permeablizing buffer (5% normal donkey serum and 0.3% Triton X-100 in PBS) for 30 min at room temperature to prevent non-specific binding. After this step, the sections were incubated overnight at 4°C with appropriate primary antibodies that were diluted in blocking/ permeablizing buffer.

The sections were washed three times (10 min each) with PBT (0.3% Triton X-100 in PBS) before incubation overnight at 4°C with appropriate fluorophore-conjugated secondary antibodies that had been diluted in PBT. Sections were washed three times in PBT (10 min each) and followed by one wash in PBS (5 min) before being mounted with anti-fade mounting gel mounting medium. To visualize nuclei, the slides were incubated with DAPI solution (25ng/ml in 1X PBS) for 10～30min before the last PBS wash.

For four-channel staining with MADM and two more markers, the staining procedure is the same as for three-channel staining, excepting for the visualization of the UV-channel signal by fluorophore Alexa Fluor 405. Alexa Fluor 405-conjugated donkey antibody was not commercially available when this work was performed. Alternatively,

we incubated the sections with Biotin conjugated species-specific donkey secondary antibody (For the UV channel) together with the fluorophore-conjugated secondary antibodies for the other three channels overnight at 4°C. After a brief wash in PBT (10 min), the sections were incubated with Alexa Fluor 405-conjugated Streptavidin (1:200) in PBT for 2 hr at room temperature. The slides were mounted in anti-fade gel mounting medium after three washes in PBT (10 min each) and one wash in PBS (5 min).

For staining including BrdU as a marker, slides used for BrdU detection were treated with HCl prior to primary antibody incubation. Briefly, after three washes in PBS, the sections were submerged in 2N HCl (in PBS) for 30 min at 37°C. The sections were then washed three times with PBS to remove HCl prior to incubation with the blocking solution.

For staining including Nestin as a marker, we found that Mouse anti-Nestin monoclonal antibody generated high background signal in mouse tissue. To remove the background signal, we used Donkey-anti-mouse monovalent Fab fragment to block the tissue sections prior to primary antibody incubation. Briefly, after incubation in blocking solution for 1 hr, the slides were washed three times in PBT (2 min each) before incubation with Fab Donkey-anti-mouse IgG (H+L) (1:10 diluted in 1X PBT) overnight at 4°C. The slides were then washed three times in PBT (10 min each) prior to incubation with anti-Nestin primary antibody.

For tissue culture staining, cells used for immunohistochemistry were cultured on poly-D-lysine coated 12mm glass coverslips. For staining, after gently removing the medium from the tissue culture plates, 0.5ml 4% PFA was added to each well and the cells were fixed at room temperature for 15 min. The coverslips were then washed one

time with 1ml PBS, dried at room temperature for 30min and were kept in 1X PBS until used for staining. To stain surface antigen O4, Triton X-100 was omitted during the entire process. Otherwise, 0.1% Triton X-100 was used for permeabilization.

Cells were incubated in blocking solution (10% NDS in PBS with or without 0.1% Triton X-100) for 1 hr at room temperature, and then incubated with primary antibody in blocking solution for 1 hr prior to three washes with PBS (with or without 0.1% Triton X-100, 10 min each). Cells were then incubated with appropriate secondary antibody (10% NDS in PBS with or without 0.1% Triton X-100) for 45 min at room temperature followed by 3 washes in PBS (with or without 0.1% Triton X-100, 10 min each). Before being mounted with anti-fade gel mounting medium, cells were incubated with DAPI for 5 min and washed one more time with PBS.

Brain regions sampled for quantification

In the present study, we define 23 distinct adult brain regions in the nonneurogenic parenchyma of sagittal brain sections for cell quantification. Selected regions cover major representative brain structures encompassing both gray and white matter, including the cerebral cortex (four random fields from rostral to caudal), the corpus callosum (four random fields from rostral to caudal), the striatum (two random fields), the ventral forebrain (ventral to striatum, two random fields), the thalamus (two random fields), the hypothalamus (two random fields), the fornix, the anterior commissure, the olfactory bulb (two random fields) and the midbrain (two random fields). Cerebellum was not included in the counting scheme because no tumor was found in this region. We also define 5 regions close to the lateral ventrical wall of adult P60 brains used for counting both slow and fast dividing cells in neurogenic SVZ, including

the dorsal lateral angle SVZ (S1) and the lateral wall SVZ (S2-S5).

Quantification of G/R Ratios from P5 to P60 mutant-MADM mouse model presented in Fig. 15

Fixed brains were cryosectioned sagittally. Three sections (20 μm in thickness) at least 200 μm apart from each other (N = 3 brains for each age) were stained and processed for data collection. To quantify MADM labeled cells and BrdU+ cells in the same brain, we stained the slides with MADM together with BrdU and DAPI. For each brain section, 9 regions were collected, including three in the cerebral cortex (Figs. 14A-14C), three in the corpus callosum (Figs. 15A, 15B and 15D), and three in the ventral side of the brain (Figs. 16A, 17A/20A, and 17B/20B). Each image was taken by a 40X objective with 1X digital zoom (0.1mm2 each image) by 2 μm optical sectioning (scanning depth 6 μm). Only the cells with clearly visible nuclei were counted.

Quantification of neurons, astrocytes, OPCs, and oligodendrocytes in nonneurogenic parenchyma of P60 mutant-MADM mouse brains presented in Fig. 16A

To count the cell number and G/R ratios of four distinct cell types in Fig. 16A, P60 brains (N = 3 brains, 4 slices from each brain) were sectioned sagittally in 14 μm (for neurons, oligodendrocytes and OPCs) or 30 μm (for astrocytes) slices. Cell types were identified by MADM-labeling and cellular markers (NeuN for neurons, CC1 for oligodendrocytes and PDGFRα for OPCs) except for astrocytes. Because Myc (to visualize RFP, which was fused with a Myc-tag at the C-terminal (Hippenmeyer et al., 2010)) and Aldolase C (for astrocytes) antibodies were both raised from goat, we omitted the Myc antibody in such staining. We found that the live color of RFP was stable for several months and strong enough in its fluorescence intensity even without antibody

staining. Therefore omitting myc staining did not affect the identification of red or double color cells.

We quantified neurons from 14 regions; and astrocytes and oligodendrocytes from 16 regions. Only cells with clearly visible cell bodies were counted. OPC quantification was performed together with BrdU long-term labeling quantification (see below). Each image was taken using a 40X objective with 1X digital zoom (0.1mm2 area) by 1 μm optical sectioning (scanning depth 10 μm). The number of red and green cells for each marker was summed together from all the regions of the four slices to calculate the average G/R ratio for the respective cell type in that brain. This value was averaged from three brains and presented as ± SEM.

Quantification of BrdU-positive OPCs in the nonneurogenic parenchyma of P60 mutant-MADM brain

Brain tissues were sectioned sagittally in 14 μm thick slices and stained with MADM plus PDGFRα (Far-red channel) and BrdU (UV channel). We collected images from 21 regions. We processed each image in two ways for different purposes. First, we counted all the PDGFRα+ cells, including red, green, yellow and non-labeled cells. This count was used to calculate the G/R ratio of OPCs and the percentage of BrdU labeling among OPCs with distinct genotypes. Second, we focused on BrdU+ cells. BrdU+ cells will be either PDGFRα+ or PDGFRα- and present one of four possible colors. To confirm that PDGFRα- BrdU+ mutant cells also belong to the OPC-oligodendrocyte lineage, we co-stained adjacent slices with MADM plus PDGFRα (far-red channel), BrdU (UV channel) and Olig2 (Red channel) (Fig. 16).

Quantification of BrdU+ cells in the neurogenic subventricular zone of P60 mutant-MADM

To identify the SVZ, we sectioned brains coronally. Four sections (30 μm in thickness, at least 200 μm apart) from each brain (three brains in total) were counted. The SVZ was divided into dorsal lateral angle SVZ (Region S1) and the lateral wall SVZ (region S2-S5). We used a 40X objective with 2X digital zoom to collect each image (imaging area is 0.025mm2 each). Bilateral SVZs in each brain section were scanned. Therefore, ten images were collected from each brain slice. The cell number from 40 images derived from a single brain was summed to calculate the G/R ratio and the percentage of mutant, WT and heterozygous cells among all BrdU positive cells in that brain. These values were averaged from N = 3 brains and presented as ± SEM.

Characterization of NG2-Cre MADM mice

NG2-Cre mediated MADM-labeling in the SVZ was analyzed in coronal cryosections (20 μm) at P360 (N = 4). Non-adjacent brain sections separated by 60 μm were collected from the rostral to caudal axis in the forebrain (i.e., collected one from every four sections), and stained with GFAP and DAPI. The SVZ was defined on each section based on DAPI staining until the SVZ region could not be clearly identified. We used a 40X objective with 2X digital zoom to scan for MADM labeled cells within the SVZ and then performed multiple optical layer confocal sectioning (0.5 μm for each step) to examine whether the MADM-labeled cells expressed GFAP.

To examine MADM-labeled cells in the olfactory bulb (OB), bilateral OBs were coronally sectioned. 100 continuous sections were collected and immunostained with anti-NeuN antibody. The granule cell layer (GCL) was identified by DAPI staining. Four

fields in the GCL of each OB section were randomly chosen using a 40X objective with 2X digital zoom. If any MADM labeled cells were found, further digital zoom (up to 4X) combined with multiple optical sectioning (0.5 μm for each step) was performed to carefully examine whether these cells co-localized with NeuN staining. 800 fields were scanned from the bilateral OBs of a single mouse (N = 4 mice).

Enrichment of OPC-like glioma cells

OPC-like tumor cells were purified from fresh glioma tissues as previously described (Cahoy et al., 2008) with minor modification. To prepare BSL-1 plates, two Petri dishes (100 × 10 mm) were incubated with BSL-1 solution (25 μg BSL-1 in 10ml DPBS for each plate). The solution was swirled carefully to ensure that each whole Petri dish was completely coated with solution prior to incubation overnight at 4°C. The next morning, the BSL-1 solution was removed from the plates and 20ml 0.2% BSA (in DPBS) was added to block the plates at room temperature until panning was performed (usually more than 4 hr).

To prepare PDGFRα immunopanning plates, two Petri dishes (100 × 10 mm) were each incubated with 10 ml of a secondary antibody solution containing Tris buffer solution (50 mM pH8.5 Tris-Cl) and 75 μg (30 μl) goat anti-rat IgG H+L chain (112-005-167) overnight at 4°C. The next morning, the secondary antibody solution was removed and the plates were washed three times with 1X DPBS. Subsequently, a primary antibody solution containing 10 μl Rat-anti-PDGFRα antibody in 10ml DPBS with 0.1% BSA was added to each plate. The plates were then incubated at room temperature until used and washed directly before immunopanning.

Cell dissociation

All dissection tools were sterilized by soaking in Cidex for 10 min and then rinsed with autoclaved distilled water prior to dissection. Tools were kept in 70% ethanol between usages. Tumor-carrying mice were euthanized by cervical dislocation, sprayed down with 70% ethanol and then wiped with Betadine to remove loose fur. The head was cut off with a pair of large dissection scissors and then soaked in 70% ethanol for 45 s to 1 min. The skin down the midline of the head was sliced through with a razor blade. The skull was cut open using a pair of small dissection scissors. Brain tissue was then lifted from the skull and washed once with sterilized cold PBS before being embedded into chilled but still liquid 4% low-melting temperature agarose (prepared in 1X PBS) on ice. After the agarose became solid, the brain was sliced into 1mm thick coronal sections and kept in cold SMEM in a Petri Dish on ice. Under a fluorescence dissection scope, tumors were easily identified because of the strong MADM fluorescence. The tumor region was dissected out with a 22.5° straight microsurgical knife. Adjacent brain slices containing some tumor tissue were preserved in 4% PFA immediately after dissection. These slices were used for immunohistochemistry and H&E staining for tumor pathology.

The remaining dissected tumor tissue was transferred to a 60 mm Petri dish containing 5ml Papain solution and diced into ～1 mm3 pieces. The Papain solution includes 1X EBSS, 0.5mM EDTA, 10mM HEPES, 26mM NaHCO3, 0.16mg/ml L-Cysteine, 20Unit/ml Papain, 1mg/ml DNase and was carefully balanced in 5% CO2 at 37°C for at least 30 min before use.

The tumor tissue was incubated in Papain solution at 37°C 5% CO2 for 90 min (gently agitated every 15 min). Afterwards it was transferred into a sterile 15ml Conical

polystyrene tube and allowed to settle to the bottom of the tube (we used a 10ml serological pipette to transfer the tissue to avoid losing the cells at this step). The Papain solution was then pipetted off and 2ml Diluted trypsin inhibitor buffer was added. Diluted trypsin inhibitor includes 1X EBSS, 10mM HEPES, 26mM NaHCO3, 1mg/ml ovomucoid, 1mg/ml BSA and was carefully balanced in 5% CO2 at 37°C for at least 30 min before use. After the tissue settled down (~2 min), the diluted trypsin inhibitor buffer was replaced by 2ml fresh diluted trypsin inhibitor buffer. The tissue was then triturated 8 times with a 5ml Pasteur pipette. After the tissue settled, 1ml of supernatant (containing dissociated cells) was carefully transferred to a new 15ml conical polystyrene tube and set aside. 1ml fresh diluted trypsin inhibitor buffer was added to the original tissue, and the trituration procedure was repeated. After the tissue settled, 1ml of supernatant was removed and combined with the other set aside supernatant. The trituration was repeated until all the diluted trypsin inhibitor buffer was used (10ml total). The last two rounds of trituration were performed with a 1ml pipette tip to ensure all the visible tissue chunks were fully dissociated into single cells.

Once the tissue was dissociated, 4ml standard trypsin inhibitor solution (10mg/ml BSA in diluted tripsin inhibitor buffer) was placed in a new 15ml conical polystyrene tube. Dissociated cell suspension was carefully layered on top of standard trypsin inhibitor solution. The layered solution was centrifuged at 220 g for 15 min followed by immediate removal of the supernatant. Cells were re-suspended in 6ml panning solution (0.02% BSA, 1mg/ml DNase in 1X DPBS) using a 5ml pipette and centrifuged at 220 g for 10 min. Cells were then re-suspended in 12 ml Panning buffer and passed through a 70 μm cell strainer before performing immunopanning.

Immunopanning

The first BSL-1 coated plate was washed 2X with PBS and 1X with DPBS. 100 µl of single cell suspension was set aside for cell number counting and cell culture. The remaining cell suspension was poured onto the first BSL-1 plate and incubated at room temperature for 15 min. During that incubation time, we counted the living cell number using the Trypan Blue method and a hemacytometer. The second BSL-1 coated plate was washed 2X with PBS and 1X with DPBS before the supernatant from the first plate was transferred to it.

During the second BSL-1 panning plate, one primary antibody coated plate was washed 3X with 20ml DPBS. The supernatant from the second BSL-1 plate was transferred to the washed primary antibody plate and incubated for 45min, and then the supernatant was transferred to the second primary antibody coated. During the second primary antibody panning plate, the first primary antibody panning plate was washed 6-8 times with DPBS to remove the non-binding cells. At the end of the last wash the plate was checked under the microscope to visualize floating cells. Additional washes were performed until there were almost zero and certainly less than 5% floating cells.

Cells were collected by incubating the first primary antibody panning plate with 2ml pre-balanced TrypLE in 5% CO2 at 37°C for 7 min. The reaction was then stopped by adding 2ml trypsin inhibitor solution (2mg/ml Ovomucoid, 0.4% BSA in 1X DPBS). Tumor cells were dislodged from the plate surface by squirting with a 1 ml pipette tip. We collected the cells and also the supernatant from the second primary antibody plate and transferred them separately to new 15ml conical polystyrene tubes which were pre-coated with 0.4% BSA at 37°C for 1 hr to prevent adherence of cells to the tubes. After

centrifugation at 220 g for 10min, the cells were washed one time with NSC medium (see cell culture section for recipes) before re-suspension in the same medium for further use (RNA extraction or allografting, see below).

WT neurosphere culture

The fetal NSCs were enriched from two independent litters of E15.5 WT embryonic brains according to the method described previously (Pollard et al., 2006) with minor modification. The forebrains (removed cerebella, olfactory bulbs and midbrains) were dissociated with Papain solution (as described above) and triturated into single cells. The dissociated cells were adjusted to 100,000 cells/ml and plated onto Poly (2-hydroxyethyl methacrylate) pre-treated 6-well tissue culture plates. The cells were cultured in NSC medium containing Neurobasal medium with B27 supplement (50X) plus Penicillin/ Streptomycin solution (100X), GlutaMAX (100X), 50ng/ml EGF and 10ng/ml FGF.

After three days of culture in 5% CO2 at 37°C, the primary neurospheres were pooled together and passaged twice as follows. We collected neurospheres in a 15ml conical polystyrene tube and let them settle for 10 min before removing the supernatant. The neurospheres were washed once with DPBS and incubated with pre-warmed 1X TrypLE for 20 min at 37°C. TrypLE was inactivated with equal volume 2X trypsin inhibitor solution (2mg/ml Ovomucoid, 0.4% BSA in DPBS) and the spheres were then dissociated into single cells by triturating 40 times with a 1ml pipette tip. Cells were centrifuged at 220xg for 10 min. The supernatant was removed and the cells were washed once with NSC medium. Cells were passed through a cell strainer and plated onto a new plate and cultured to form secondary neurospheres. Three to four days later secondary

neurospheres were dissociated into single cells as above. Dissociated cells were expanded in NSC medium using 1X adherent 25cm2 PDL-coated flask for 4 days and passaged one more time in 3X 25cm2 PDL-coated flasks. The expanded cells were trypsinized off the flasks, washed, and frozen in 15% DMSO (in NSC medium) in liquid nitrogen for further use or directly lysed in TRIzol for RNA analysis.

Real-time quantitative PCR

Total RNA was extracted by TRIzol followed by phenol/chloroform extraction and ethanol precipitation (Miller et al., 2009). cDNA was then synthesized using Superscript III and Oligo (dT) Primers. qRT-PCR primers were designed using MacVector v10.0 and Primer 3 Plus to cover two exons flanking a > 500bp intronic region. Gene amplification was tested by at least two distinct functional primer sets. Only primer sets with the best performance for each gene were used for further analysis.

qPCR was done using an ABI 7900HT Real-Time PCR System for 40 cycles by denaturing at 95°C for 15 s, annealing at 58°C for 30 s and elongating at 72°C for 30 s, followed by the default dissociation curve program. PCR amplification was performed in RT2 Real-Time SYBR Green/Rox PCR Master Mix. Relative levels of cDNA were measured by using control primers for β-Actin (ActB) and/or Glyceraldehyde-3-Phosphate Dehydrogenase (Gapdh) to normalize gene expression for each sample. CT values were measured within the geometric amplification phase and averaged for triplicate reactions. Reactions with a CT value of 35 or higher were considered null. Values for individual genes were normalized by subtracting the average of the internal control gene(s) Gapdh and/or ActB.

Microarray and gene expression analysis

44K Mouse Development Oligo Microarrays from Agilent Technologies were used for microarray analysis. For each experiment, total RNA was fluorescently labeled and hybridized directly against a common reference sample generated from the RNA pool of four WT P17 mouse brain neocortexs. After scanning the images by using a GenePix 4000B scanner, the intensity values were extracted by using the Agilent Feature Extraction Software. In this study, we only focused on the channel of the tumor samples to extract their intensity values, but ignored the reference channel signals. Probe sequences were downloaded from the Agilent website and aligned to a transcript sequence database consisting of 49,040 Mus musculus genes from Ensembl57. After removing probes that did not align, or aligned with more than 5 mismatches (n = 19,194), or probes that aligned to multiple genes (n = 2,283), 23,006 probes representing 17,585 genes remained. Expression profiles were quantile normalized and log transformed. Expression values for genes represented by multiple probes were collapsed by taking the mean value of the set of probes.

Data from Cahoy et al. (2008) were downloaded and pre-processed as described previously (Verhaak et al., 2010). The data set contains 38 mouse samples, which clustered into 8 clusters after selecting 2,000 variably expressed genes. Gene sets were defined by selecting the 250 genes with the highest absolute t test statistic value, except for the OPC cluster with only two cluster members, where fold change values were used. The clusters used in the present study contained the following samples: (1) Postnatal day 16 myelin oligodendrocytes (n = 4); (2) Postnatal day 6 to 7 OPCs (n = 2); (3) Postnatal day 7 neurons (n = 4);. (4) Postnatal day 17-30 astrocytes (n = 6). Gene sets for four

previously described GBM subtypes (Verhaak et al., 2010) were defined by selecting the top 250 marker genes as provided, and by mapping the human genes to mouse genes using Ensembl57.

Single sample GSEA has been described elsewhere (Barbie et al., 2009; Verhaak et al., 2010). In short, genes were ranked by their expression values. The empirical cumulative distribution functions (ECDF) of both the genes in the signature as well as the remaining genes were calculated. An enrichment score was obtained by a sum of the difference between a weighted ECDF of the genes in the signature and the ECDF of the remaining genes. This calculation was repeated for all signatures and samples. A positive score indicates a high ranking of up-genes in the signature, and low ranking of down-genes in the signature. A negative value does not indicate the opposite, but rather a lack of effect.

Results

Establishment of an MADM-based genetic mosaic model for glioma

MADM-based cancer modeling needs three prerequisites: cancer-causing gene mutation(s), MADM cassettes that reside centromerically to the prospective mutated genes, and a Cre transgene that expresses in a certain tissue or organ. To establish an MADM-based glioma model, we decided to inactivate *p53* and *NF1*, both of which are among the most frequently mutated genes in human glioma patients (McLendon et al., 2008; Parsons et al., 2008) and have been used to model glioma in mice (Reilly et al., 2000; Zhu et al., 2005). We also engineered MADM cassettes into the *Hipp11* genomic locus (Fig. 14A) proximal to the *p53* and *NF1* chromosomal locations on mouse Chr. 11

(Hippenmeyer et al., 2010). To induce MADM-based recombination, we chose *hGFAP*- or *Nestin-Cre* transgenes (termed *NSC-Cre* hereafter) that are expressed in both embryonic and adult NSCs (Petersen et al., 2002; Zhuo et al., 2001), thereby generating MADM-labeled cells in all NSCs-derived lineages: neurons, astrocytes, oligodendrocytes, and OPCs (Fig. 14B). With the ability to identify MADM-labeled cell types, we can readily analyze the aberrant growth of GFP+ mutant cells by directly comparing to their RFP+ WT counterparts within each lineage in the same mouse brain.

MADM-based glioma model offers the opportunity to trace the entire tumorigenic process

After recombining mutant alleles of *p53* and *NF1* with MADM alleles to generate a mutant-MADM mouse model, we analyzed overall expansion of green mutant cells at different ages. In contrast to the comparable distribution of green and red cells in WT-MADM brains in which all labeled cells are wild-type (Fig. 15A, left), we observed a progressive overrepresentation of green mutant cells in mutant-MADM brains from postnatal day 5 (P5) to P60 (Fig. 15A, top three of the right column) and eventually the formation of GFP+ full-blown tumors at ~5 months of age (Fig. 15A, bottom-right).

Because mutant (green) and wild-type (red) sibling cells originate from the same mother cell in equal numbers initially (Fig. 14A), the ratio of green to red cell numbers (referred to as the G/R ratio hereafter) allows us to quantitatively evaluate the extent of mutant cell expansion. A G/R ratio equal or close to 1 indicates no growth advantage of the mutant population, whereas a G/R ratio > 1 indicates a growth advantage of mutant over WT cells (referred to as "overexpansion" hereafter). From P5 to P60, the average G/R ratio in mutant-MADM brains increased from ~4 to ~30 (Fig. 15B), indicating a

continuous overexpansion of mutant cells. However, the increase of G/R ratio appeared to reach a plateau between P30 and P60, suggesting that overexpansion of mutant cells had largely stopped between those ages. In addition to quantifying cell numbers, we also compared the proliferative status between mutant and WT cells with a brief BrdU pulse to label cells undergoing DNA replication. At all time points analyzed, the percentage of BrdU+ cells in the mutant population was significantly higher than that in the WT population (Fig. 15C). Furthermore, whereas WT cells largely ceased to proliferate at early postnatal age (P10), some mutant cells (∼0.3% among all mutant cells population) continued to divide and contributed to the entire population of BrdU+ cells in the brain parenchyma at P60 (Fig. 15D). Thus, mutant cells showed not only elevated, but also prolonged proliferative capacity compared to WT cells at a stage before tumor formation (termed the "pretransforming" stage hereafter).

At later ages (4–5 months), all mutant-MADM mice developed brain tumors with strong Ki67 staining (Fig. 15A, bottom-right). Tumors were invariably GFP+ (n = 28), indicating that they originated from MADM-induced mutant cells. Although transcriptome profiles of these tumors were very similar (see below), pathological analysis revealed great heterogeneity in their appearance: some showed typical astrocytic features; some showed malignant glioma features such as necrosis, multinuclear giant cells, and perivascular and perineuronal satellitosis; and most were highly anaplastic (data not shown, Figs. 17B–17D, and Fig. 18D). Taken together, MADM allows us to trace mutant cells at all gliomagenic stages and provides the analytical accessibility between initial mutations and the final transformation.

Mutant OPCs manifest dramatic overexpansion in pretransforming MADM brains

We next sought to determine the identity of glioma cell of origin by separately determining the G/R ratio of NSCs and of each NSC-derived cell type at P60, when mutant cell expansion had largely ceased but tumors were yet to arise. We predicted that, at this ostensibly dormant stage in tumorigenesis, the mutant cell type capable of transformation should manifest significant overexpansion and maintain sustained proliferative activity. We first quantified the G/R ratios of all four NSC-derived cell types (Fig. 16A). The average G/R ratio from different brain regions showed slight reduction of mutant neurons and only minor expansion of mutant astrocytes and oligodendrocytes. In stark contrast, the G/R ratio of OPCs was > 130, significantly higher than those of the other three cell types (Fig. 16A and data not shown). Notably, mutant OPCs appeared to have impaired differentiation potentials to give rise to mature oligodendrocytes, as the G/R ratio of oligodendrocytes was ~10 times less than that of OPCs (Fig. 16A). Moreover, brief BrdU pulses at P60 exclusively labeled mutant OPCs in brain parenchyma, although the total number was low, suggesting that mutant OPCs were progressing but were still relatively dormant at this age (Fig. 16B, left). Remarkably, despite such a dramatic overrepresentation of mutant OPCs, we did not observe any pathological features in these brains by conventional pathological analysis (data not shown), demonstrating the capability of MADM to probe into a previously inaccessible stage during tumor development.

Mutant OPCs constitute the majority of the proliferation pool in pretransforming MADM brains

It seems paradoxical that we found nearly full penetrance of tumor formation in

our model at 4–5 months of age, yet dividing cells at P60 were so rare (Fig. 15C and Fig. 16B, left). Considering that pretransforming cells could undergo a prolonged cell cycle, we evaluated the slow-dividing cell population by extending the duration of BrdU administration and observed that 30% of all mutant OPCs were able to incorporate BrdU (Fig. 16B, right, and data not shown), suggesting that many mutant OPCs were dividing slowly. We noticed that a small fraction of BrdU-positive mutant cells in the brain parenchyma did not express the OPC marker PDGFRα (Figs. 16B, right, and 16C). To determine whether these were cells from other lineages or oligodendrocytes differentiated from recently divided OPCs, we costained brain sections adjacent to those shown in Fig. 16C with antibodies against BrdU, PDGFRα, and Olig2, a "pan-oligo" marker that expresses in both OPCs and oligodendrocytes. Using a "4+1 channel" staining scheme (Figs. 16C–16E), we found that all BrdU+, PDGFRα-negative cells examined expressed Olig2 (Fig. 16D), suggesting that they were newly differentiated oligodendrocytes from initially BrdU-labeled mutant OPCs. Taken together, our detailed analysis reveals that, among all mutant cell types, OPCs are the only proliferative population in the brain parenchyma at pretransforming stages.

Although we did not observe the same extent of overexpansion in other mutant cell types as that in mutant OPCs at P60, it could be a result of increased cell death in neuronal and astrocytic lineage following the overexpansion of NSCs. Therefore, it is critical to directly analyze the proliferative activity of adult NSCs within the SVZ (Doetsch et al., 1999). After 7 days of BrdU feeding of mutant-MADM mice, we quantified all BrdU+ cells in the SVZ, which include type B, C, and A cells (Doetsch et al., 1999). In stark contrast to our findings in the brain parenchyma, where 85% of BrdU-

positive cells were mutant (Fig. 16G, left), only ~1% of BrdU-positive cells were mutant in the SVZ (Fig. 16G, right). More importantly, the G/R ratio of BrdU+ cells in adult SVZ was not significantly different from 1 ($p = 0.39$), indicating that there was no overexpansion of mutant adult NSCs (Fig. 16F). Furthermore, we did not observe any overexpansion of mutant granule neurons in the olfactory bulb, which are the main progeny generated by adult NSCs (Fig. 16F). Therefore, *p53* and *NF1* mutations seemed unable to enhance the proliferative activity of adult NSCs. Taken together, detailed comparison between mutant and WT cell behaviors in NSCs and all progeny lineages strongly suggests that NSCs function as the cell of mutation but fail to directly transform, whereas OPCs function as the cell of origin for glioma.

Tumor cells in fully developed gliomas exhibit salient OPC features

Following the investigation at pretransforming stages, we next analyzed the expression pattern of cell-specific markers in malignant tumors generated in the MADM-glioma model (Figs. 17A–17D). Under low magnification, we observed prominent enrichment of NSC markers Nestin and Sox2 and the astrocytic marker GFAP, but not neuronal or oligodendrocyte markers in the tumor region (Fig. 17E). This observation is consistent with previous findings in both mouse models and human patients (Alcantara Llaguno et al., 2009; Louis et al., 2007; Wang et al., 2009; Zhu et al., 2005) and has been considered as evidence to support the NSC origin of gliomas. However, it is important to note that these tumors also showed enriched expression of OPC markers, such as Olig2, PDGFRα, NG2, CD9, and O4 (Fig. 17E and data not shown). The expression of some OPC markers was further confirmed by quantitative RT-PCR from crude tumor samples (Fig. 17F).

To clarify whether these markers were expressed by tumor or bystander cells, we analyzed the tumor sections at higher magnification and found that GFP+ proliferating tumor cells expressed many markers for OPCs (Figs. 17G-I), but not for neurons or astrocytes (Figs. 17J and 17K). In some tumors, there were a few GFP+ cells expressing oligodendrocyte marker CC-1, but not Ki67 (Fig. 17L), which could be either residential cells or oligodendrocytes differentiated from tumor cells. To extend the molecular characterization of the cellular identity, we next performed global transcriptome comparison between tumor samples and all four neuroglial cell types (Cahoy et al., 2008) and confirmed that tumor cells closely resembled OPCs, but not any other neuroglial cell types (Fig. 17M). Importantly, regardless of diverse pathological features, all tumor samples shared almost identical molecular profiles (Fig. 17M and data not shown), suggesting a common cell of origin for these morphologically heterogeneous tumors.

In line with immunofluorescent staining (Fig. 17E), we observed elevated Nestin level in some, but not all, tumors by transcriptome and qRT-PCR analyses (data not shown). However, it is unclear whether Nestin is expressed in OPC-like tumor cells or in bystander cells. To clarify this problem, we enriched for OPC-like tumor cells by using an immunopanning method widely used to purify WT OPCs based on their surface expression of PDGFRα (Cahoy et al., 2008). The transcriptome of purified OPC-like tumor cells displayed an even higher extent of similarity to that of normal OPCs (Fig. 17M), demonstrating that these cells contributed to the OPC signatures of glioma samples. qRT-PCR results showed that, in addition to consistent expression of OPC markers, purified OPC-like cells from some tumors expressed Nestin at a comparable level to NSC-derived neurospheres (data not shown). This finding was further confirmed

by marker costaining of the tumor mass, showing that some GFP+ dividing tumor cells coexpressed both PDGFRα and Nestin (data not shown). Importantly, regardless of their Nestin expression levels, these panned tumor cells, when orthotopically allografted into NOD-SCID mouse brains, effectively initiated secondary tumors (data not shown) that recapitulated both histological (data not shown) and molecular features (Figs. 17F and 17M) of their primary tumors. Taken together, these data demonstrate that highly proliferating tumor cells manifested salient OPC features. Although Nestin was expressed in some tumor cells, the fact that primary tumor cells with either undetectable or high levels of Nestin expression could efficiently initiate new tumors implies that Nestin expression appears irrelevant to tumorigenicity and is likely acquired during transformation. Therefore, analyses of tumor cells further support OPCs, rather than NSCs, neurons, or astrocytes, as the cell of origin for glioma in *p53*, *NF1* mutation-driven glioma model.

Localization of earliest neoplastic lesions suggests a gray matter origin of glioma

The analyses of lesion locations at early stages of transformation could provide great insights for understanding the cancer cell of origin. We first analyzed morphological or cyto-architectural changes of mutant cells in fully developed tumors, reasoning that some features could serve as landmarks of transformation, thereby helping to determine the locations of early lesions. In all malignant tumors, even in the smallest lesions identifiable by classic pathology, MADM staining revealed a prominent cyto-architectural feature in which multiple glioma cells wrapped around a single neuronal cell body (Fig. 18C), a salient pathological feature known as perineuronal satellitosis in human glioma (Fig. 18D). It should be noted that normal OPCs are often juxtaposed to

neuron cell bodies (data not shown), implicating that perineuronal satellitosis in tumors might stem from this unique cyto-architecture of OPCs.

When we compared OPC::neuron association from pretransforming to tumor-bearing brains, we found that the number of mutant OPCs for each accompanying neuron changes from mostly one (occasionally two) prior to transformation (Figs. 18A and 18E) to frequently three or higher in any Ki67-positive lesions (Figs. 18B, 18C, and 18E). Therefore, we used "three or more mutant OPCs per neuron" as a landmark to examine a cohort of brains for initiating lesions that cover a period from right after the dormant stage to around the time when tumors can readily be identified (Figs. 18F–18K). In 2.5- to 3.5-month old mice, we observed the earliest detectable lesions in cortical gray matters (Fig. 18F, P75–P90 and top of P105 column). In further developed lesions, tumor cells were more frequently found to overlap with white matter tracks and the SVZ (Fig. 18F, bottom of P105, P120, and > P120 columns). This observation indicates that previous findings of glioma in white matter tracks and the SVZ (Persson et al., 2010; Zhu et al., 2005) might be the path or endpoint, rather than the starting point, of tumor cell migration and further argues against NSCs and for OPCs as the cell of origin.

OPCs can be directly transformed into malignant glioma by p53/NF1 mutations

Our results thus far point toward the following scenario for glioma development in our model: NSCs carrying *p53*/*NF1* mutations give rise to mutant OPCs, which multiply, progress, and eventually transform into malignancies. However, it remains unclear whether mutations must occur in NSCs or whether OPCs can be directly transformed by the same set of mutations. To address this question, we next used *NG2-Cre* that has been reported to specifically express in OPCs, but not NSCs (Komitova et

al., 2009; Zhu et al., 2008), to perform MADM analysis. We first verified that *NG2-Cre* exclusively labels OPCs and oligodendrocytes, but not mature astrocytes or neurons in brain parenchyma of WT-MADM mice (Figs. 19A–19D). Next, we examined the entire SVZ of WT-MADM mice and never found any GFP- or RFP-labeled cells coexpressing GFAP (data not shown). Lastly, we thoroughly examined the olfactory bulbs (OB) from four WT-MADM mice and did not find any NeuN+ MADM-labeled green or red cells (data not shown), further supporting the absence of *NG2-Cre* expression in NSCs.

In stark contrast to the extremely sparse labeling in WT-MADM mouse brains induced by *NG2-Cre* (Fig. 19E), the green mutant OPCs in mutant-MADM mice populated the entire brain (Fig. 19F) and reached an average G/R ratio of more than 300 (Fig. 19G) at P60. Consistent with findings from the NSC-Cre-based MADM model, the differentiation process of mutant OPCs was hampered (Figs. 19G and 19H). These data demonstrate that introduction of *NF1* and *p53* gene mutations directly into OPCs is sufficient to drive overproliferation and to impair OPC differentiation.

After 8 months of age, we found GFP+ tumors in brains of *NG2-Cre*-induced mutant-MADM mice at nearly full penetrance (Fig. 19I, Fig. 19J, and data not shown). These tumors showed marker staining and transcriptome profiles indistinguishable from *NSC-Cre*-induced malignant gliomas (Figs. 17F, Fig. 17M, and data not shown). Dissociated cells from these tumors could effectively initiate secondary tumors in allografting assays (data not shown), demonstrating the malignancy of these OPC-initiated tumors.

Despite their indistinguishable marker expression patterns, transcriptome profiles, and tumor-initiating ability, we observed that tumors induced by *NSC-Cre* tended to

reside in the brain parenchyma, whereas those induced by *NG2-Cre* were more frequently found around the ventral hypothalamic region and often invaded into the subarachnoid space (Fig. 19I, Fig. 20B, and data not shown). However, tumors induced by the two *Cre* lines at equivalent locations showed indistinguishable pathological features (Figs. 20A and 20B). Moreover, although tumor cells with subarachnoid invasion were morphologically distinct from those in the brain parenchyma, transplanting them into the brain parenchyma of NOD-SCID mice resulted in secondary tumors that were morphologically indistinguishable from those primary tumors in the brain parenchyma (Fig. 20C). These data strongly suggest that the variation in tumor pathology is merely an effect of tumor location. The differences in latency and spatial distribution between *NG2-Cre*- and *NSC-Cre*-induced tumors most likely are attributed to technical reasons, such as the much lower expression level of NG2-Cre transgene or possible spatial biases of these *Cre* lines. It is also possible that these two models have distinct noncell-autonomous cues for tumor formation. For instance, *NSC-Cre* generates heterozygous astrocytes or neurons, whereas *NG2-Cre* does not, which might function as special niches to facilitate tumor formation and to bias the tumor location. Nevertheless, all lines of evidences, including marker expression patterns (Figs. 17F–17L and data not shown), transcriptome profiles (Fig. 17M), tumorigenic capacity, and pathology (Fig. 20 and data not shown), indicate that tumors from both models are intrinsically identical. Therefore, our data demonstrate that OPCs can be directly transformed into malignant glioma as a result of concurrent *p53*/*NF1* mutations.

Finally, we investigated the human relevance of our studies by comparing the transcriptome profile of tumor cells in our model with the molecular signatures of human

glioma samples. Among four subtypes of human malignant glioma revealed by recent studies from the Cancer Genome Atlas (TCGA) (Verhaak et al., 2010), both *NSC-Cre*- and *NG2-Cre*-initiated glioma models match well with the proneural subtype (Fig. 17M), which has the poorest responses to chemo- and radiotherapies (Verhaak et al., 2010). Therapeutic strategies designed according to the intrinsic OPC nature should have promising effectiveness for proneural or even other subtypes of glioma.

Discussion

In this study, we used the mouse genetic mosaic system termed MADM to track the entire tumorigenic process, with the aim of identifying the cell of origin for glioma when initial concurrent mutations of *p53* and *NF1* occur in embryonic NSCs. Taking advantage of the in vivo cellular resolution and WT internal control cells afforded by MADM, we analyzed lineage-specific cellular aberrations at premalignant stages and observed dramatic overexpansion and elevated proliferative activity in the OPC lineage but minimal abnormalities in NSCs and all other progeny lineages. Furthermore, marker staining, cyto-architecture, and transcriptome analyses all point to OPCs as the transforming cell type for malignant glioma. Finally, direct introduction of *p53*/*NF1* mutations into OPCs led to the formation of glioma that were indistinguishable from those initiated from NSCs, based on both transcriptome and pathological analyses. Therefore, our studies clarified that, for the glioma model induced by *p53* and *NF1*, though NSCs could serve as the cell of mutation, OPCs serve as the actual cell of origin.

MADM as a tool to distinguish cancer cell of origin and cell of mutation

The identification of the cancer cell of origin is crucial but often controversial,

even with the use of genetic modified mouse models (Visvader, 2011). Conceptually, caution must be taken that cells acquiring initial mutations (cell of mutation) may not directly transform into malignancy because their transforming potential could be manifested by their progeny cell types. For example, when *Patched* mutation was introduced into NSCs in a mouse model of medulloblastoma, malignant transformation only occurred after NSCs gave rise to lineage-restricted granular neuron precursors (GNPs) (Schüller et al., 2008; Yang et al., 2008). In this scenario, the transforming potential of NSCs (cell of mutation) is only manifested by mutant progeny GNPs, the cell of origin that possesses the conducive signaling context. Therefore, to unambiguously identify cancer cell of origin, one has to genetically dissect the transforming potentials of all progeny derived from the initial mutated cell type. This could be feasible if highly specific *Cre*-transgenic lines are available, such as in the case of medulloblastoma studies (Schüller et al., 2008; Yang et al., 2008). However, such a strategy cannot be generalized to other cancer models due to the lack of specific *Cre*-transgenic mice in every lineage for most tissues. An alternative approach is to analyze aberrant growth in all cell lineages derived from initial mutant cells during pretransforming stages (Visvader, 2011), which is now possible with

MADM-based cancer models

Our study extends rather than contradicts previous reports. Limited by the cellular resolution at pretransforming stages, previous studies demonstrated the transforming potential of NSCs but did not clearly distinguish cell of mutation from cell of origin (Alcantara Llaguno et al., 2009; Wang et al., 2009; Zhu et al., 2005). MADM, on the other hand, offers a robust analytical paradigm for the identification of the cell of origin

with both permanently GFP-labeled mutant cells and RFP-labeled WT internal control cells immediately after the initial mutations occurred. In principle, MADM could be used for all types of cancer, as long as oncogenic mutations for a particular type of cancer, a tissue-specific stem cell Cre-transgenic line, and cellular markers of each lineage within that tissue are available.

Our studies emphasize the importance of intersection between genetic mutations and the signaling context within the cell of origin. We show here that OPCs are particularly sensitive to *p53*/*NF1* mutations, whereas NSCs and other brain cell types are much less responsive. Interestingly, it has been reported that the *NF1* mutation alone can promote OPCs to proliferate in mice and zebrafish (Bennett et al., 2003; Lee et al., 2010), suggesting that *NF1* plays a pivotal role in regulating OPC proliferation with conserved mechanisms across species. Such intersections should be further exploited to help understanding of the molecular mechanism of OPC transformation, which should provide critical insights for developing effective therapeutic strategies. Notably, whereas OPCs appear to be particularly responsive to *NF1* mutation, other cell types, including NSCs, could be specifically vulnerable to other genetic lesions. It would be very interesting to investigate whether different mutations could transform distinct cell of origins, which might account for the great heterogeneity of human glioma.

OPC is an important glioma cell of origin with underappreciated proliferative and plastic potentials

One of the notions that support the prevalent view of NSCs as the glioma cell of origin is the persistent proliferative activity of NSCs in the brain during the entire life, which makes them susceptible to oncogenic mutations. Additionally, NSCs share many

cellular properties with glioma cancer stem cells, such as the capability to self-renew, the potential to differentiate into multiple cell lineages, and the expression of some common cellular markers such as Nestin and Sox2. However, a large body of work has revealed that many of these features are also shared by OPCs. First, OPCs in fact represent the largest proliferative pool in the brain parenchyma for both rodents and human (Dawson et al., 2003; Geha et al., 2010). Second, OPCs at early developmental stages have been found to express many commonly used stem cell markers, including Nestin and Sox2. Third, OPCs isolated from rat optic nerve can be readily reprogrammed into a multipotent NSC-like status under in vitro conditions, which become self-renewable and can differentiate into astrocytes and neurons in addition to oligodendrocytes (Kondo and Raff, 2000). Intriguingly, OPCs from human subcortical white matter intrinsically behave like NSCs without reprogramming (Nunes et al., 2003), raising the possibility that primate OPCs are more plastic than their rodent counterparts. Taken together, the intrinsic nature of OPCs renders them great susceptibility to oncogenic mutations, which could be harnessed to devise effective therapeutic strategies.

Previously, it has been reported that the injection of PDGF-BB-expressing virus into the corpus callosum, the expression of PDGF-BB under CNPase promoter, and the overexpression of v-erbB under human S100β promoter all led to glioma formation with OPC features (Assanah et al., 2006; Lindberg et al., 2009; Persson et al., 2010). Although providing circumstantial evidence, these studies have faced challenges to unequivocally pinpoint OPCs as the cell of origin for malignant glioma (Visvader, 2011) due to noncell-autonomous effects caused by the secretion of PDGF-BB (Assanah et al., 2006; Lindberg et al., 2009) and the use of non-OPC-specific promoter to drive the initial tumorigenic

event (Persson et al., 2010). By circumventing all of these caveats, our studies now firmly establish that OPCs serve as the cell of origin for malignant glioma with relevant genetic mutations. For clinical applications, intrinsic OPC properties should be fully exploited to design effective therapies for glioma, especially for the proneural subtype. For example, understanding the unique proliferative capacity of OPCs could help devise treatments to stall tumor progression; deciphering the migration mechanism of OPCs could enhance the effectiveness of surgery by preventing the infiltration of tumor cells into the entire brain; and probing into the differentiation process of OPCs could facilitate the design of differentiation therapy strategies.

The application of MADM beyond cancer cell of origin studies

Compared to conventional models, the MADM-based cancer model bears significant advantages for exploring critical cancer biology problems beyond the cell of origin. Its reliable labeling of tumor cells at the single-cell resolution should help to distinguish bona fide tumor cells from bystander cells in a complex tumor mass, which is highly valuable for dissecting tumor architecture, tracking metastasized tumor cells, and studying tumor-niche interactions. Importantly, the availability of a built-in internal control enables quantitative comparison between mutant and WT sibling cells. Such rigorous phenotypic analyses within the same animal largely remove variations often introduced by comparing phenotypes between individual animals, thereby helping to identify subtle but important phenotypes that could be easily missed by using conventional approaches.

Using *Drosophila* genetic mosaic system, combined with the application of positive marking with gene manipulation at the mosaic level, helped to elucidate many

fundamental questions in cancer biology, such as identifying important oncogenes and TSGs by forward genetic screening (Potter et al., 2001; Tapon et al., 2001) and deciphering the complicated tumorigenic processes such as metastasis (Pagliarini and Xu, 2003). Along this same principle, we expect that a further modified MADM system could address even more sophisticated problems in cancer biology beyond the capability of the current one. For instance, by replacing fluorescent reporters with a transcription factor such as tTA or rtTA to specifically manipulate gene expression in mosaic mutant cells, one could model "second hit" events or therapeutic strategies and gain deep insights into the molecular network for cancer formation. By incorporating novel magnetic resonance imaging reporters such as H-ferritin (Genove et al., 2005) into MADM cassettes, one could noninvasively trace tumor initiation, progression and metastasis in real time. Also, by combining with transposon-based mutagenesis such as *Sleeping Beauty* (Collier et al., 2005) or *PiggyBac* (Ding et al., 2005), the MADM system should be able to perform forward genetic screening for novel recessive TSGs in mouse somatic cells without involving a large number of animals.

Concluding remarks

In summary, our study demonstrates the importance of analyzing the entire process of tumor development for identifying cancer cell of origin. MADM-based tumor modeling can be applied in principle to any other tumor type and should help to resolve many important problems in cancer biology. Broadly, the ability to perform sporadic single-cell genetic manipulations in unambiguously labeled cells should make MADM an invaluable analytical tool for fields such as developmental biology and neuroscience that rely heavily on in vivo analyses.

CHAPTER V

TU-TAGGING: CELL TYPE-SPECIFIC RNA ISOLATION FROM INTACT COMPLEX TISSUES

This work was published in the journal *Nature Methods* in June 2009. Myself and Kristin Robinson performed the microarray-based and immunohistochemistry expression analyses. Kristin Robinson and Michael Cleary performed the northern-blot analysis. Chris Doe was the principle investigators for this work.

Introduction

Cell type–specific gene expression is a defining feature of multicellular organisms (Tomancak et al., 2007). The analysis of cell type–specific transcriptomes can provide insight into the mechanisms used to generate cellular diversity (Wang et al., 2009) as well as help determine the underlying cause of disease (Schadt et al., 2003). Although a few methods are available for cell type–specific RNA isolation (Roy et al., 2002; Tanke and van der Keur, 1993; Doyle et al., 2008; Heiman et al., 2008) each has constraints, and researchers are often limited by their ability to isolate RNA from cell types of interest (Nelson et al., 2006). Thus, developing new methods for cell type–specific RNA isolation is an important goal for genomic analysis of development and disease.

We previously had shown that the *Toxoplasma gondii* nucleotide salvage enzyme uracil phosphoribosyltransferase (UPRT) can be used to biosynthetically label newly synthesized RNA *in vivo* (Cleary et al., 2005). Under natural conditions, UPRT couples ribose-5-phosphate to the N1 nitrogen of uracil to yield uridine monophosphate (UMP), which is subsequently incorporated into RNA. When the modified uracil analog 4-thiouracil is provided to UPRT as a substrate, the resultant product is also incorporated

into RNA, and this incorporation has little effect of cellular physiology (Cleary et al., 2005). Thio-substituted nucleotides are not a natural component of nucleic acids, and the resulting thio-labeled RNA can be readily tagged and purified using commercially available reagents. Owing to the ability to use this method to separate newly synthesized RNA from bulk cellular RNA, we and others have used it to measure RNA synthesis and decay rates (Cleary et al., 2005; Dolken et al., 2008).

Here we describe a different use for 4-thiouracil UPRT-based biosynthetic labeling that we call 'TU-tagging'. We reasoned that by spatially restricting UPRT expression in a multicellular organism, 4-thiouracil will be modified and subsequently incorporated into newly synthesized RNA only in cells expressing UPRT. Thus, even if RNA is isolated from the whole organism, RNA from the cells expressing UPRT can be recovered by purifying labeled RNA (Fig. 21A). This method would be particularly useful for isolating RNA from cell types that are difficult to isolate by dissection or dissociation methods, such as subsets of neurons or glia in the central nervous system.

Methods

UPRT and 4-thiouracil toxicity tests

We crossed *UAS-UPRT* lines to various GAL4 lines to express UPRT in several cell types (see below) and ubiquitously (using *tubulin-GAL4*). In no case did we observe lethality, developmental defects or delay in the time to pupariation and eclosion (data not shown). We next monitored development of larvae ubiquitously expressing UPRT when grown on food containing 4-thiouracil (4TU). We observed no effect on central nervous system development (developmental timing or neuroanatomy) in larvae grown on 4TU

for 48 h or less; we then used an 8 h or less feeding interval for all experiments. Longer exposure to 4TU (with or without UPRT expression) produced slight developmental delays but did not affect viability.

Fly stocks

We used standard methods to clone the *T. gondii* UPRT coding sequence into pUAST to generate the *UAS-UPRT* and *UAS-HA-UPRT* plasmids, and obtained independent viable insertions on the X, II or III chromosome. Three *UAS-UPRT* lines consistently had low background expression and high GAL4-induced expression: *UAS-HA:UPRT2.1* (chromosome II), *UAS-UPRT3.1* (chromosome III) and *UAS-HA:UPRT3.2* (chromosome III). We crossed one of these three lines to the following previously described *GAL4* lines for all experiments (chromosome in parentheses) (i) *worniu-GAL4(II)* × *UAS-HA:UPRT3.2*; (ii) *OK107-GAL4(IV)* × *UAS-HA:UPRT3.2*; (iii) *c179-GAL4(II)* × *UAS-HA:UPRT3.2*; (iv) *elav-GAL4(III)* × *UAS-HA:UPRT3.2*; (v) *tubulin-GAL4(III)* × *UAS-HA-UPRT3.2*; (vi) *prospero-GAL4(III)* × *UAS-UPRT3.1*; (vii) *twist-GAL4(II)*;*how24B-GAL4(III)* × *UAS-UPRT3.1*; and (viii) *repo-GAL4(III)* × *UAS-HA:UPRT3.2*.

4TU treatment and RNA extraction

To treat embryos with 4TU, the embryos were dechorionated in bleach, washed, rinsed with isopropanol, blotted dry and submerged in octane for 3 min (all in the basket). Embryos were blotted and air dried for ~3 min until soft to touch. Embryos were transferred to Schneider's media containing 1.0 mM 4TU for 2 h at 25 °C or 30 °C, blotted dry, moved using a paintbrush to an Eppendorf tube containing 1 × PBS with 1% tween (PT) and centrifuged at 6,000g for 30 s. PT was removed and the embryos were

homogenized in Trizol and stored at −80 °C until RNA purification. To treat larvae with 4TU, larvae of the desired stage were placed on mocha food caps (20 ml H20, 0.4 g sucrose, 0.18 g agar, 1 g dextrose and 0.5 g brewers yeast) containing 0.5 mM 4TU for the indicated time at 30 °C, homogenized in Trizol and stored at −80 °C until RNA purification. To treat adult flies with 4TU, they were starved for ~16 h, placed on mocha food containing 1.0 mM 4TU for 6–8 h at 25 °C, homogenized in Trizol and stored at −80 °C until RNA purification. A more detailed protocol is available on request.

Total RNA was extracted from Trizol using standard methods with the following additional steps: an initial centrifugation at 12,000g for 10 min at 4 °C to remove insoluble material followed by a 5-min incubation at room temperature (20–25 °C) to ensure complete dissociation of RNA-bound proteins. Only RNA samples with absorbance 260/280 ratios of ≥ 2.0 were used for subsequent biotinylation and purification steps. In all cases, RNA samples were resuspended at a final concentration of ≥0.4 μg μl−1.

Purification of TU-tagged RNA

Detailed methods for biotinylating and purifying thio-tagged RNAs have been published (Cleary et al., 2005; Zeiner et al., 2008), and detailed protocols including the most recent improvements are available upon request. Relevant changes to previous protocols are summarized here. Biotinylation of RNA was performed using EZ-Link biotin-HPDP, as previously described. Biotinylation reactions contained 10 mM Tris (pH 7.4), 1 mM EDTA and 2 μl of a 1 mg ml−1 biotin-HPDP solution (in dimethylformamide) per 2 μg of RNA. The reaction volume was adjusted with water so that the concentration of biotin-HPDP was equal to 30% of the final reaction volume.

Biotinylation reactions were incubated in the dark for 3 h at 25 °C before RNA precipitation. Biotinylated RNA was detected by blotting and probing with streptavidin-horseradish peroxidase as previously described. Purification of biotinylated TU-tagged RNA was performed as previously described with the following modifications: 2 μl of MPG streptavidin beads were used per microgram of input RNA. The input RNA was always at a concentration of 0.5 μg μl−1. After blocking with yeast tRNA and washing, beads were resuspended in the input RNA sample plus a volume of MPG buffer (1 M NaCl, 10 mM EDTA and 100 mM Tris (pH 7.4) in RNase-free H2O) equal to one-third the input RNA volume. Beads plus RNA were incubated at room temperature for 20 min before collecting the non-bound sample and washing with MPG buffer (one 5-min wash at room temperature, two 1-min washes at room temperature, one 1-min wash in 65 °C MPG buffer and one final 1-min wash at room temperature). After the removal of as much MPG buffer as possible, TU-tagged RNA was eluted by incubating the beads for 10 min in freshly prepared 5% 2-mercaptoethanol. RNA was precipitated using isopropanol and linear acrylamide. After resuspending RNA in water, samples were placed in the magnetic stand again to remove any remaining MPG beads. For the purification procedures, input amounts of biotinylated RNA were 14–20 μg.

Microarray analysis

We used the Agilent eArray platform to design a custom *D. melanogaster* oligonucleotide microarray representing 14,141 unique genes from the Flybase release 5.4 genome. Fifty nanograms of TU-tagged and 200 ng untagged RNA were fluorescently labeled and hybridized directly against each other to Agilent microarrays. The microarray experiments were performed according to Agilent's protocol (Version 5.5, February 2007)

and scanned using an Axon GenePix 4000B scanner. Fluorescent ratios for each microarray element were recovered and normalized using GenePix Pro 6.0.

Normalizing for transcript uracil number

We downloaded the dmel-all-transcript-r5.4.fasta file from Flybase, which contained the sequences of every predicted transcript from genome release 5.4. We wrote a Perl script to count the number of uracils in a transcript (transcript uracil number) for each transcript. In cases where there was alternative splicing, we averaged the uracil number over the multiple isoforms. This was necessary because, for the most part, our microarray did not distinguish between different isoforms of the same gene. To normalize the data, these uracil counts were plotted against the observed microarray ratios using OpenOffice.org Spreadsheet and the regression equation was determined. This equation was used to calculate the expected ratio for each transcript based on the transcript's uracil number. For each transcript, the normalized ratio was calculated by subtracting the expected ratio from the observed ratio. After plotting the initial normalized ratios, we noticed a group of transcripts, all with low uracil numbers, that had very high normalized ratios. Upon investigation we noticed that these transcripts were missing annotations for either or both untranslated regions. Thus, our transcript uracil counts for these transcripts were lower than their actual uracil number, leading to incorrect normalization. Therefore, transcripts with missing untranslated region annotations were excluded from normalization and subsequent analysis. We note that normalization errors owing to transcript misannotation will decrease as transcriptome annotations improve.

Results

To test the ability to biosynthetically label RNA in *Drosophila*, we delivered 4-thiouracil and monitored its incorporation into RNA by purifying total RNA, performing thio-biotin coupling and using streptavidin–horseradish peroxidase to detect labeled RNA (Methods). Wild-type larvae or adult flies, or larvae or adult flies containing only the GAL-4–inducible transgene UAS-UPRT (upstream activating sequence–UPRT) but no source of GAL4, had very small amounts of labeled RNA when fed 4-thiouracil (Fig. 21B). In contrast, 4-thiouracil–fed larvae or adult flies containing both *GAL4* and *UAS-UPRT* transgenes expressed UPRT in the cell types containing GAL4 (data not shown) and showed robust RNA labeling (Fig. 21B). Similarly, embryos soaked in 4-thiouracil–containing medium had robust RNA labeling only when both *UAS-UPRT* and *GAL4* were present (Fig. 21C). We conclude that the combination of UPRT and 4-thiouracil can be used to biosynthetically label RNA in *Drosophila* embryos, larvae and adults.

To determine the limits of sensitivity, we fed larvae 4-thiouracil and expressed UPRT in different subsets of the larval brain. When UPRT was expressed in about 2,000 neurons in the entire larva, we detected small amounts of labeled RNA (Fig. 21B). When we reduced the number of UPRT-expressing cells to about 250 neural progenitors per larva, we observed no detectable signal over background (Fig. 21D). However, a simple dissection of the intact larval brain before RNA purification yielded excellent signal and dramatically reduced background RNA labeling (Fig. 21D). We conclude that the amount of TU-tagged RNA correlated well with the number of cells expressing UPRT; that simple tissue isolation can reduce background and increase sensitivity; and that non–central nervous system tissue contributed to nearly all of the low-level background RNA

labeling in larvae.

To confirm that thio-labeled RNA was from UPRT-expressing cells and demonstrate the utility of TU-tagging for cell type–specific RNA isolation, we purified TU-tagged and untagged RNA and compared them using microarrays. We isolated RNA from larval glia, which are present in low numbers, are highly dispersed and have a complex cell morphology (Fig. 22A), making them one of the most difficult cell types to isolate by dissection or dissociation methods. We used reversed polarity (*repo*)-*GAL4* to drive expression of *UAS-UPRT* specifically in glial cells of the larval brain (Fig. 22A). We purified TU-tagged and untagged RNA from 72–96-h larval brains and hybridized them to custom Agilent microarrays (Methods). We detected signal for 7,354 of the 14,141 genes present on the microarray. If TU-tagged RNA was primarily from UPRT-expressing cells, known larval glia–specific genes should be among the most enriched in the TU-tagged relative to untagged RNA. There are four genes known to be expressed specifically in larval glia but not neurons or trachea of the brain. All four were enriched greater than twofold in the TU-tagged RNA (Fig. 22B) and three of the four were among the 5% most enriched genes *ana* (3.54-fold, top 0.3%, 19/7,354), *repo* (2.85-fold, top 1.6%, 115/7,354) and *moody* (2.75-fold, top 1.9%, 141/7,354), which are expressed in all larval glia (Bainton et al., 2005; Ebens et al., 1993), and *ced-6* (2.13-fold, top 7.6%, 561/7,354), which is expressed in a subset of larval glia (Awasaki et al., 2006). Thus, known larval glia–specific genes were enriched in the TU-tagged RNA. We conclude that TU-tagging can effectively isolate glia-specific RNA from whole brain tissue without prior cell dissociation.

We next tested whether the number of uracils in a transcript influenced the extent

of enrichment we observed, because long transcripts containing many uracils are expected to be labeled at a higher frequency than short transcripts with few uracils. Thus we plotted the number of uracils in a transcript against the observed microarray ratio for each transcript (Methods). There was a notable positive correlation (Fig. 22C). Transcripts with many uracils were often among the most enriched transcripts overall, even when they were not enriched relative to other transcripts with a similar number of uracils (Fig. 22C), whereas transcripts with few uracils that were clearly enriched relative to other transcripts with a similar number of uracils were unlikely to be among the most enriched transcripts overall (Fig. 22C). We conclude that the number of uracils in a transcript is a source of bias in TU-tagging experiments.

To remove this bias on the microarray results, we used the regression equation to calculate the expected TU-tagged to untagged ratio and subtracted it from the observed ratio for each transcript (Methods). This normalization procedure successfully removed the bias in data that resulted from the number of uracils in a transcript (Fig. 22D) and had a large impact on which genes were present in the top 1% or 5% most enriched genes (Fig. 22E). We conclude that enrichment bias owing to the number of uracils in a transcript can be removed using a simple normalization procedure and that normalization has a large impact on which genes are considered most enriched.

We next determined how the normalization affected the enrichment of the known larval glia–specific genes. Without normalization three of four previously known larval glia–specific genes were within the top 5% of enriched genes (Fig. 22C). After normalization all four were within the top 3.2% of enriched genes: *ana*, top 0.2% (12/6,167); *repo*, top 1.6% (99/6,167); *moody*, top 2.4% (150/6,167); and *ced-6*, top 3.2%

(235/6,167) (Fig. 22D). Although the characterization of new glia genes was beyond the scope of this work, these results suggest that the other highly enriched genes in this dataset are excellent candidates for regulating aspects of larval glia biology. We conclude that normalizing for transcript uracil number improves TU-tagging data analysis and that normalization should be useful for other 4-thiouracil/UPRT–based methods.

Discussion

An important property of the TU-tagging method is that only newly synthesized RNAs are labeled. Thus, the percentage of labeled cellular RNA will depend on the duration of labeling. Although long labeling periods should work well for isolating the majority of RNA present in a particular cell type, short labeling periods could be used to detect changes in gene expression at successive time points in specific cell types, because newly synthesized RNA could be separated from bulk cellular RNA. This would be useful for studying rapid changes in gene expression after a particular developmental or physiological event.

The TU-tagging method is likely to work well in other systems including vertebrates. There is very little UPRT-independent 'background' incorporation of 4-thiouracil into RNA in mouse or human cell culture lines (Cleary et al., 2005; Rinn et al., 2008). Spatial control of UPRT in vertebrates could be achieved by using transgenes, as in *Drosophila*, or by retroviral delivery, electroporation or mRNA injection. Thus, at least *Drosophila*, mice and humans appear suitable for TU-tagging experiments. It is likely that TU-tagging can be used for cell type–specific RNA isolation in many multicellular organisms and should be particularly useful for the study of development, neurobiology

and disease.

APPENDIX A

TABLES

Table 1. Candidate genes for down-stream targets of *sox9a* and *sox9b* in zebrafish.

Protein type			Expression					Ref
ID	**Gene name**	**Chr**	**PA**	**Eye**	**Brain**	**Others**	**Fig.**	
Transcription factor								
BF938407	*crx*	5		+	+		Figs. 2, 5	1, 2
AF025348	*vsx1*	17		+			Figs. 2, 4, 5	3, 4
AF036148	*neurod*	9		+	+		Figs. 2, 5	5, 6, 7
BI840999	*sox4a*	19		+	+		Figs. 2, 5	8, 9, 10
AF392995	*klf2b*	2				PF	Fig. 7	11, 12
Cell–cell interaction								
BG308285	*rs1*	N.A.		+			Figs. 2, 4, 5	10
Extra cellular matrix								
AF192763	*col11a2*	19	+				Fig. 1	N.A.
U23822	*col2a1a*	8	+				Fig. 5	13, 14
Others								
BM070961	*calb2a*	7		+	+		Figs. 3–5	9, 10
AW077148	*calb2b*	18		+	+		Figs. 3–5	9, 10
Unknown EST								
AI722369	N.A.	N.A.				PF, M	Fig. 7	N.A.

Chr, Chromosome location; PA, Pharyngeal arch; Ref, References; PF, Pectoral fin; M, Muscle; N.A., not available. 1, (Liu et al., 2001); 2, (Shen and Raymond, 2004); 3, (Passini et al., 1997); 4, (Passini et al., 1998); 5, (Korzh et al., 1998); 6, (Mueller and Wullimann, 2002); 7, (Ochocinska and Hitchcock, 2007); 8, (Mavropoulos et al., 2005); 9, (Rauch et al., 2003); 10, (Thisse and Thisse, 2004); 11, (Oates et al., 2001); 12, (Thisse et al., 2001); 13, (Yan et al., 1995); 14, (Yan et al., 2005).

Table 2. Mutants affecting brain neuroblast numbers used in this study.

Genotype	Synonym	type I / II neuroblast phenotype	References
lgl^{334}	*lgl*	ectopic type II	(Bowman et al., 2008)
$brat^{11}$	*brat*	ectopic type II	(Bowman et al., 2008)
wor-gal4 UAS-aPKC^{CAAX}	*aPKC*^{CAAX}	ectopic type I (some II)	(Bowman et al., 2008)
$aurA^{8839}$	*aur*	ectopic type I (some II)	(Bowman et al., 2008)
lgl^{-} lgd^{d7}	*lgl lgd*	ectopic type II	this work
lgl^{334}; $pins^{62}$	*lgl pins*	ectopic type I (some II)	this work

Table 3. Representation of cell type-specific genes within microarray groups A, B, and C.

Gene	Cell type (in larval brain)	Array group	References
wor	neuroblast	A	(Ashraf et al., 2004)
dpn	neuroblast	A	(Bier et al., 1992)
cycE	neuroblast	A	(Caldwell and Datta, 1998)
grh	neuroblast	A	(Uv et al., 1997)
dmyc	neuroblast	A	(Betschinger et al., 2006)
E(spl)mγ	neuroblast	A	(Almeida and Bray, 2005)
insc	neuroblast	A	(Parmentier et al., 2000)
mira	neuroblast	A	(Peng et al., 2000)
pros	type I neuroblast, GMC, neurons	B	(Boone and Doe, 2008)
ase	type I neuroblast, GMC	B	(Bowman et al., 2008)
elav	neurons	C	(Robinow and White, 1988)
gcm	glia	C	(Hosoya et al., 1995)
repo	glia	C	(Xiong et al., 1994)

APPENDIX B

FIGURES

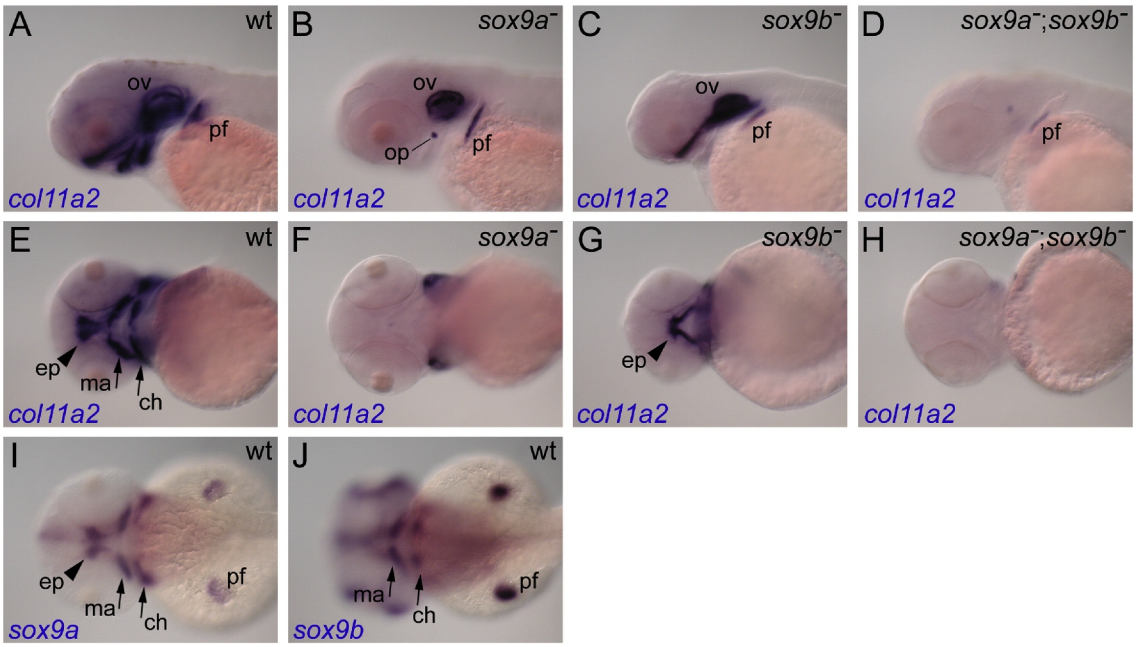

Fig. 1. Differential regulation of *col11a2* expression by *sox9a* and *sox9b*. (A–D) Lateral view and (E–H) ventral view of 2 dpf embryos. Expression of *col11a2* appeared in neurocranium (ethmoid plate, arrowhead), pharyngeal arches (mandibular and ceratohyal, arrows), otic vesicle and in the pectoral fin bud (A, E). Expression in neurocranium and pharyngeal arches was lost in the *sox9a* mutant (B, F). In the *sox9b* mutant, expression disappeared from the pharyngeal cartilages, but expression in the neurocranium (arrowhead) remained (C, G). Expression almost disappeared in *sox9a*;*sox9b* double mutant embryos (D, H). At 2dpf, *sox9a* was expressed in the neurocranium (arrowhead) and pharyngeal arches (arrows) in locations corresponding to *col11a2* expression (I), but *sox9b* was only expressed in mandibular and ceratohyal (arrows), not in the ethmoid plate and trabeculae (J). These results confirmed microarray results showing that *col11a2* is down-regulated in *sox9* mutant embryos, and revealed that *sox9a* and *sox9b* regulate *col11a2* expression in a tissue specific manner according to *sox9* expression patterns. ch, ceratohyal; ep, ethmoid plate; ma, mandibular; ov, otic vesicle; op, opercle; pf, pectoral fin bud.

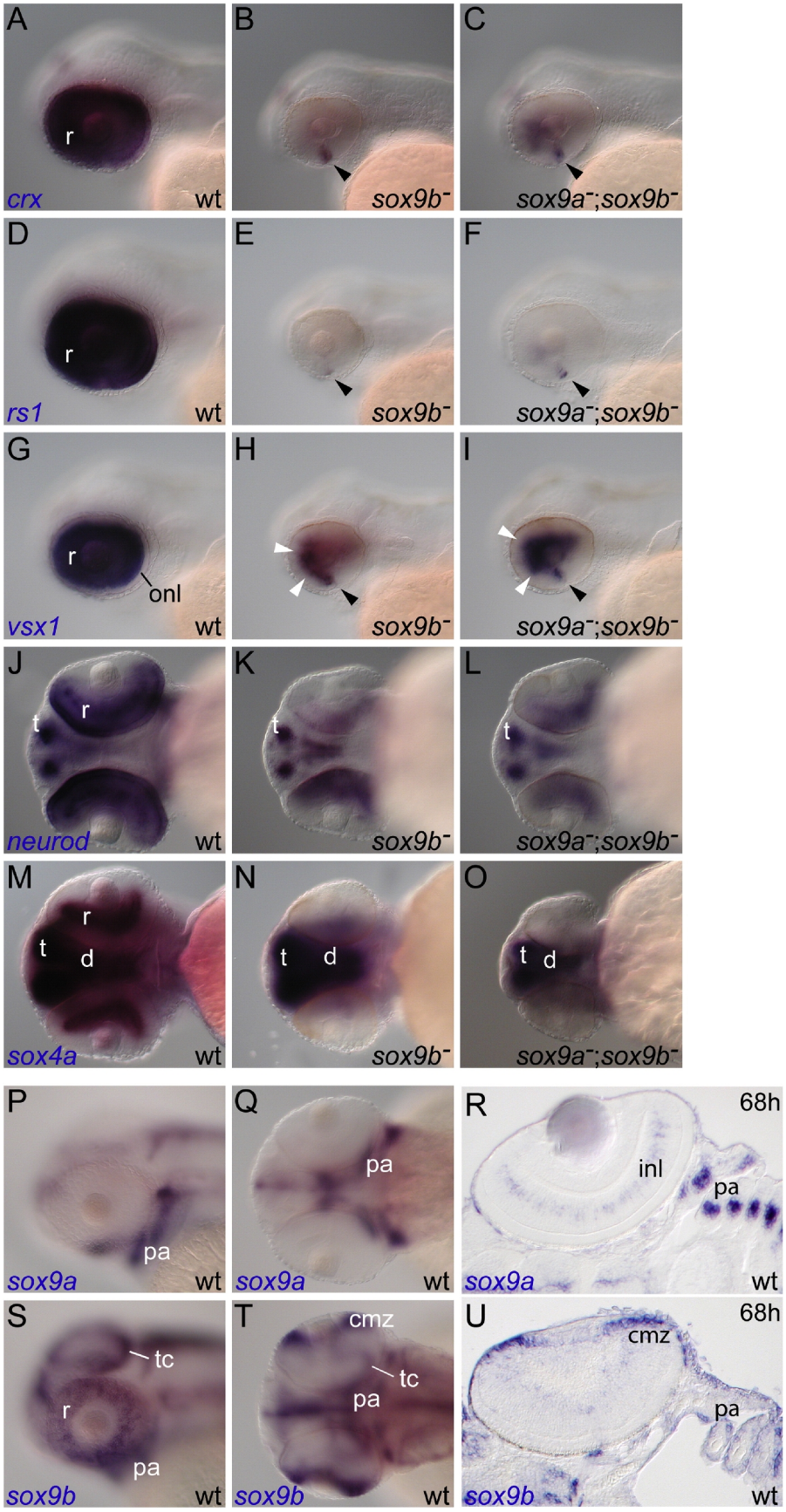
A
r
crx
wt
B
sox9b-
C
sox9a-;sox9b-
D
r
rs1
wt
E
sox9b-
F
sox9a-;sox9b-
G
r
onl
vsx1
wt
H
sox9b-
I
sox9a-;sox9b-
J
t
r
neurod
wt
K
t
sox9b-
L
t
sox9a-;sox9b-
M
r
t
d
sox4a
wt
N
t
d
sox9b-
O
t
d
sox9a-;sox9b-
P
pa
sox9a
wt
Q
pa
sox9a
wt
R
68h
inl
pa
sox9a
wt
S
tc
r
pa
sox9b
wt
T
cmz
tc
pa
sox9b
wt
U
68h
cmz
pa
sox9b
wt

Fig. 2. Expression of candidate *sox9* down-stream targets analyzed by in situ hybridization analysis in 2 dpf embryos, the same stage as used for microarray analysis. Anterior to the left. Lateral views of *crx* (A–C), *rs1* (D–F) and *vsx1* (G–I) show that they are strongly expressed in developing retina; *crx* and *rs1* but not *vsx1* were expressed in the outer nuclear layer. Expression was severely reduced in *sox9b* mutants and double mutants; only a small ventral-nasal patch of retinal expression (black arrowhead) was detected for *crx* (B, C) and *rs1* (E, F), and ventro-nasal expression in the retina (white arrowhead) was observed for *vsx1* (H, I). Ventral view of *neurod* (J–L) and *sox4a* (M–O) showing *neurod* expression in the outer nuclear layer and inner nuclear layer of the retina and telencephalon (J) and *sox4a* expression in inner part of retina, telencephalon and diencephalon (M). Retinal expression of *neurod* was reduced in *sox9b* and double mutant embryos, but the expression in the telencephalon was not affected in these mutants, demonstrating specificity (K, L). Expression of *sox4a* in the retina was lost in *sox9b* mutants and double mutants, whereas expression in telencephalon and diencephalon was not affected in the mutants (N, O). At 2 dpf, *sox9b* (S, T), but not *sox9a* (P, Q), was expressed in retina, consistent with the reduced gene expression pattern shown above. Lateral view (P, S) and ventral view (Q, T). At 68 hpf, *sox9a* is expressed in the inner nuclear layer (R), and *sox9b* is expressed in ciliary marginal zone (U). These results confirmed the microarray result that *crx*, *rs1*, *vsx1*, *neurod* and *sox4a* are down-regulated in *sox9b* mutant embryos, and revealed that *sox9b* is required for expression of these candidate genes in retina at 2 dpf. cmz, ciliary marginal zone; d, diencephalon; onl, outer nuclear layer; pa, pharyngeal arches; r, retina; t, telencephalon; tc, tectum.

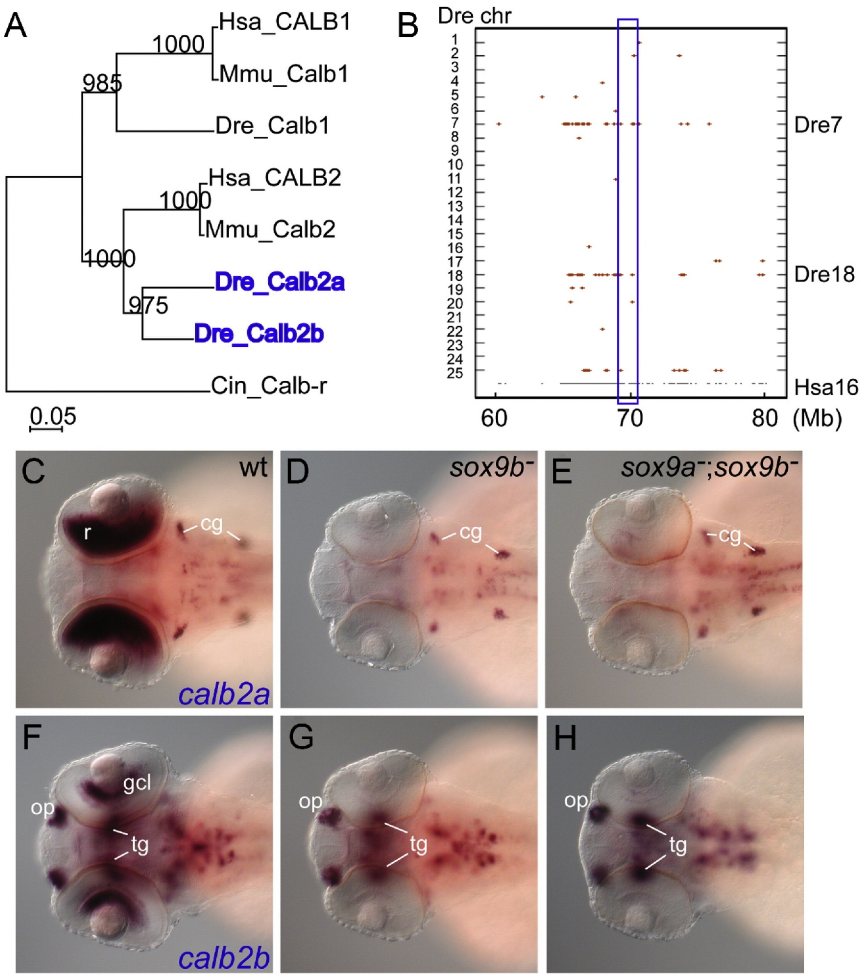

Fig. 3. Candidate target genes *calb2a* and *calb2b* are co-orthologs of tetrapod *Calb2* and their expression was lost in the *sox9b* mutant retina. (A) Phylogenetic analysis of Calbindin family proteins using the neighbor-joining method. Numbers on the branches are bootstrap values for the group in a thousand runs. Zebrafish Calb2a and Calb2b branched together as co-orthologs of tetrapod Calb2 with very high bootstrap support. Ascidian Calbindin-related protein was used as an outgroup. Cin: ascidian, *Ciona intestinalis*; Dre: zebrafish, *Danio rerio*; Hsa: human, *Homo sapiens*; Mmu: mouse, *Mus musculus*. (B) A conserved synteny around Human *CALB2* on *Homo sapiens* chromosome 16 (Hsa16) with zebrafish *calb2a* on *Danio rerio* chromosome 7 (Dre7) and *calb2b* on Dre18. Red dots represent zebrafish genes orthologous to human genes in the position 60–80 Mb of Hsa16. Dorsal view of *calb2a* (C–E) and *calb2b* (F–H) expression. Expression of *calb2a* appeared in the retina, cranial ganglia and hindbrain neurons (C). Expression in the retina was not observed in *sox9b* mutants (D) and double mutants (E), but other expression domains were not affected, demonstrating specificity (D, E). The expression of *calb2b* was observed in the olfactory placode, tegmentum and ganglion cell layer (F). Expression in the retina was not detected in *sox9b* mutants (G) and double mutants (H), but expression in the other tissues was not affected (G, H). These results showed that zebrafish *calb2a* and *calb2b* are co-orthologs of the vertebrate *Calb2* gene, that they are differently expressed in retina, and that both genes depend on *sox9b*

function. cg, cranial ganglia; gcl, ganglion cell layer; op, olfactory placode; r, retina; tg, tegmentum.

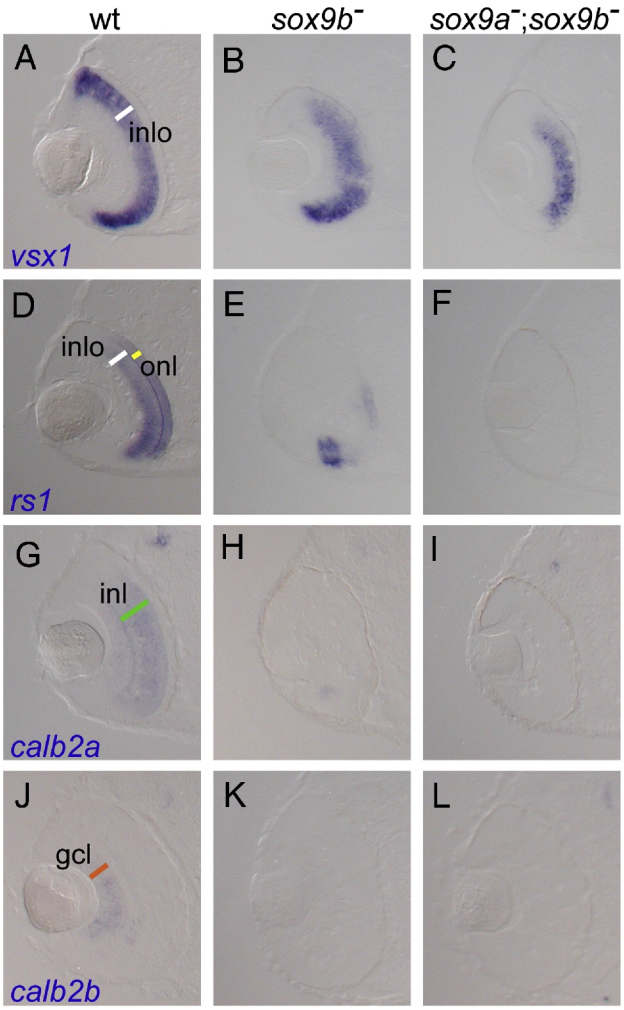

Fig. 4. Candidate genes were expressed in different layers in the developing retina. (A–C), *vsx1*; (D–F), *rs1*; (H–I), *calb2a*; (J–L), *calb2b*. Expression of *vsx1* was observed in the outer part of the inner nuclear layer (A). Expression was reduced especially in the dorsal part in *sox9b* mutants (B) and double mutants (C). In wild-type embryos, *rs1* was expressed in the outer part of the inner nuclear layer and in the outer nuclear layer (D). In *sox9b* mutants and *sox9a*;*sox9b* double mutant embryos, expression was severely reduced or nearly gone except for the ventral ciliary marginal zone (E, F). Expression of *calb2a* was observed in the inner nuclear layer (G) and *calb2b* was expressed in the ganglion cell layer (J). In each case, expression was severely reduced in *sox9b* mutants (H, K) and *sox9a*;*sox9b* double mutant embryos (I, L). White bar in (A, D), inlo, outer part of inner nuclear layer; yellow bar in (D), onl, outer nuclear layer; green bar in (G), inl, inner nuclear layer; orange bar in (J), gcl, ganglion cell layer.

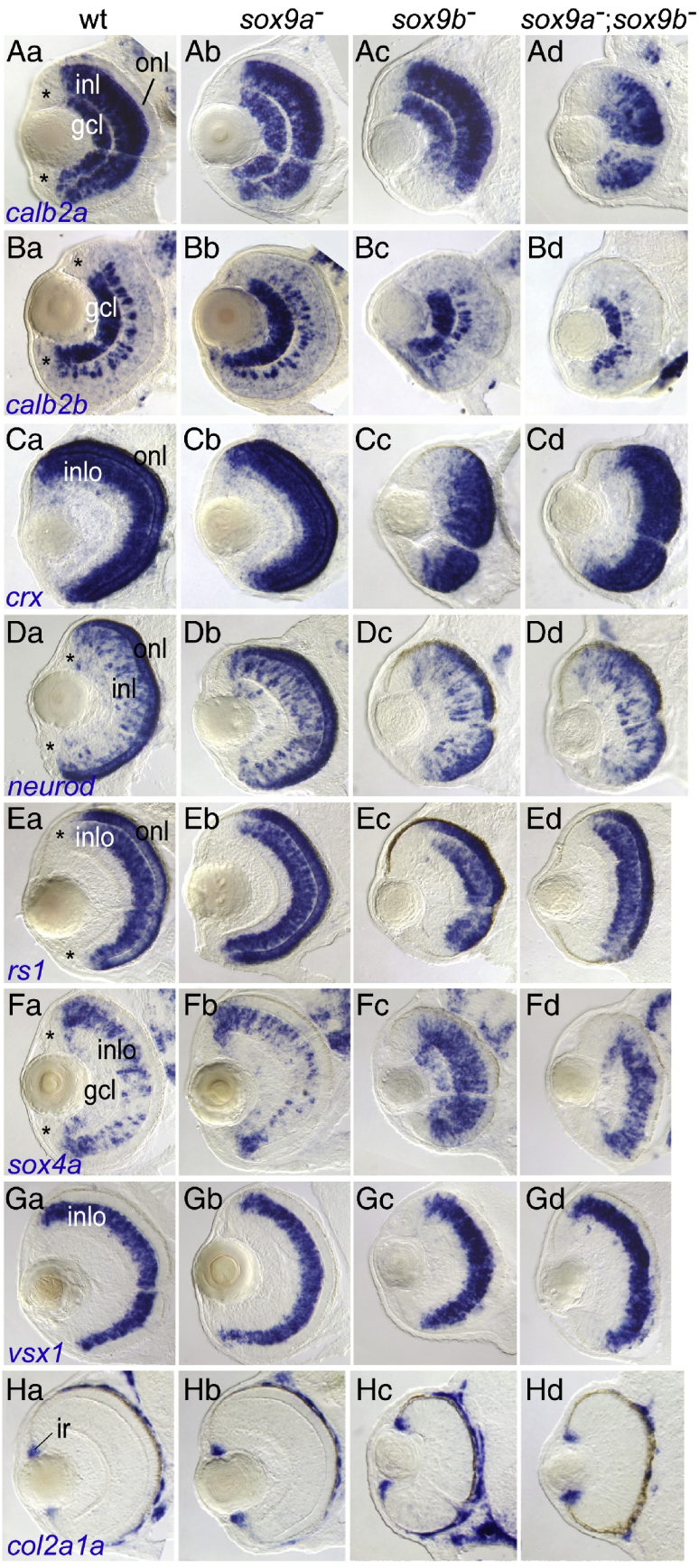
wt
sox9a-
sox9b-
sox9a-;sox9b-
Aa
onl
inl
gcl
calb2a
Ab
Ac
Ad
Ba
gcl
calb2b
Bb
Bc
Bd
Ca
inlo
onl
crx
Cb
Cc
Cd
Da
onl
inl
neurod
Db
Dc
Dd
Ea
inlo
onl
rs1
Eb
Ec
Ed
Fa
inlo
gcl
sox4a
Fb
Fc
Fd
Ga
inlo
vsx1
Gb
Gc
Gd
Ha
ir
col2a1a
Hb
Hc
Hd

Fig. 5. Expression of candidate genes in retina at 3 dpf. In situ hybridization on histological sections with *calb2a* (Aa, wild-type; Ab, *sox9a* mutant; Ac, *sox9b* mutant; Ad, double mutant), *calb2b* (Ba–Bd), *crx* (Ca–Cd), *neurod* (Da–Dd), *rs1* (Ea–Ed), *sox4a* (Fa–Fd), *vsx1* (Ga–Gd) and *col2a1a* (Ha–Hd). Expression of *calb2a* appeared in the ganglion cell layer and inner nuclear layer (Aa); *calb2b* was expressed in the ganglion cell layer and a cell population in the inner part of the inner nuclear layer (Ba); *crx* was expressed in the outer part of the inner nuclear layer and outer nuclear layer (Ca); *neurod* was expressed in the outer nuclear layer and a cell population in the inner nuclear layer (Da); *rs1* was expressed in the outer part of the inner nuclear layer and the outer nuclear layer (Ea); *sox4a* was expressed in a cell population in the inner nuclear layer (Fa); and *vsx1* was expressed in the outer part of the inner nuclear layer (Ga). The ciliary marginal zone did not express *calb2a*, *calb2b*, *neurod*, *rs1* or *sox4a* (asterisk). In *sox9b* and double mutants, expression in dorsal margin and ventral margin was reduced for all seven genes. The expression of *sox4a* in the inner nuclear layer was stronger in *sox9b* and *sox9a*;*sox9b* double mutants than in wild-type and *sox9a* mutants (Fa–Fd). Expression of *col2a1a* appeared in iris (Ha), and was not affected in *sox9* mutant embryos (Hb–Hd). No obvious defect was observed in the *sox9a* mutant retina. cmz, ciliary marginal zone; gcl, ganglion cell layer; inl, inner nuclear layer; inli, inner part of inner nuclear layer; inlo, outer part of inner nuclear layer; ir, iris; onl, outer nuclear layer.

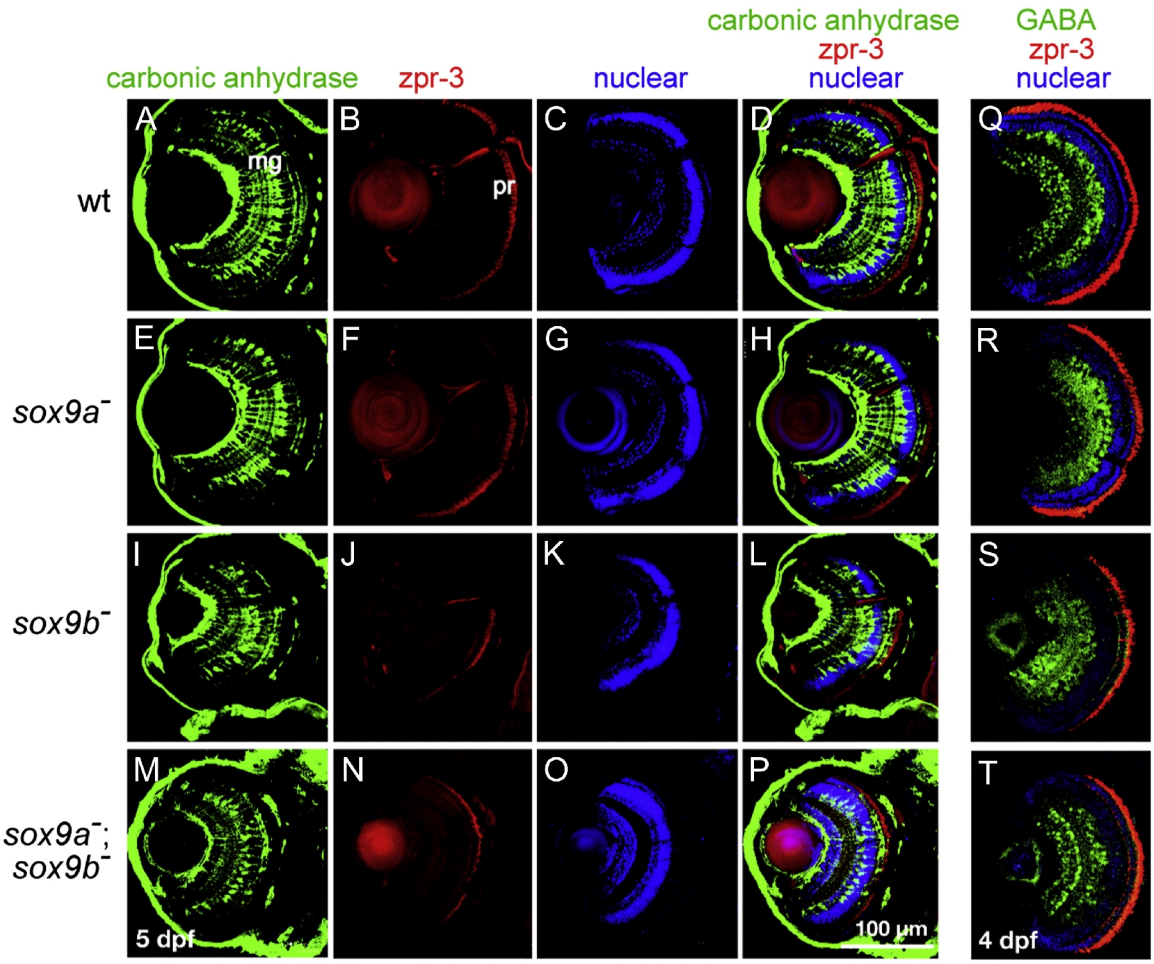

Fig. 6. Fluorescent antibody staining of retinal cell types in wild-type and *sox9* single and double mutant larvae. (A–P) 5 dpf and (Q–T) 4 dpf. Eye morphology in *sox9a* mutants was generally similar to that in wild-type siblings. In contrast, *sox9b* mutants and *sox9a*;*sox9b* double mutants had smaller eyes and eye morphology was less organized than in wild-type siblings and *sox9a* mutants. (A, E, I, M) Retinal Müller glial cells labeled by anti-carbonic anhydrase showed that *sox9b* mutants (I) and *sox9a*;*sox9b* double mutants (M) had fewer Müller glial cells compared to wild-types (A) and *sox9a* (E) siblings. (B, F, J, N) Photoreceptor cells labeled by zpr-3 showed that fewer and thinner photoreceptor cells were present in *sox9b* mutants (J) and *sox9a*;*sox9b* double mutants (N) compared to wild-type (B) and *sox9a* (F) siblings. (C, G, K, O) Nuclear layers of the retina stained by nuclear dye (TO-PRO-3 iodide) showed that all the nuclear layers were present in *sox9* mutants. (D, H, L, P) green: anti-carbonic anhydrase, red: zpr-3, and blue: nuclear staining. (Q–T) Retinal amacrine cells labeled by anti-GABA: green, photoreceptor cell labeled by zpr-3: red, and nuclear layers of the retina staining by nuclear dye (TO-PRO-3 iodide): blue. The results showed that *sox9a* mutants (R), *sox9b* mutants (S) and *sox9a*;*sox9b* double mutants (T) have fewer organized amacrine cells and photoreceptor cells compared to wild-type siblings (Q). Scale bar: 100 µm.

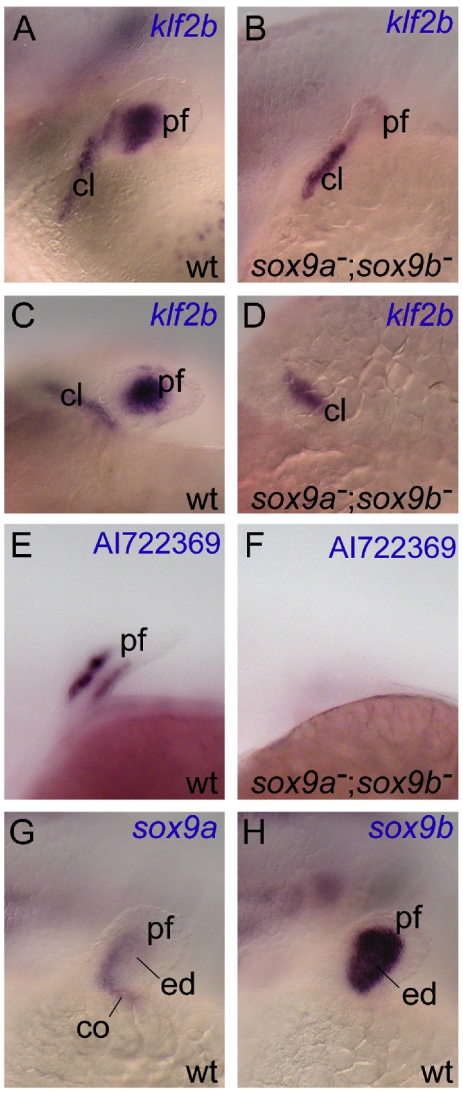

Fig. 7. Expression of candidate genes in the pectoral fin bud. (A, B) Lateral view and (C, D) dorsal view of *klf2b* expression. In wild-type embryos, *klf2b* was expressed in the mesenchyme of the pectoral fin bud and in the adjacent cleithrum (A, C). Expression in the pectoral fin bud was not detected in double mutants (B, D). Expression of *klf2b* in the cleithrum was not affected in the mutant. Expression of EST AI722369 was observed in the presumptive muscle in the pectoral fin bud (E). AI722369 expression did not appear in *sox9a*;*sox9b* double mutant embryos (F). Lateral view of *sox9a* (G) and *sox9b* (H) expression. At this stage, *sox9a* was expressed in an L-shape in the pectoral fin bud (G), and *sox9b* expression appeared more broadly in the pectoral fin bud (H). cl, cleithrum; co, scapulocoracoid; ed, endochondral disc; pf, pectoral fin bud.

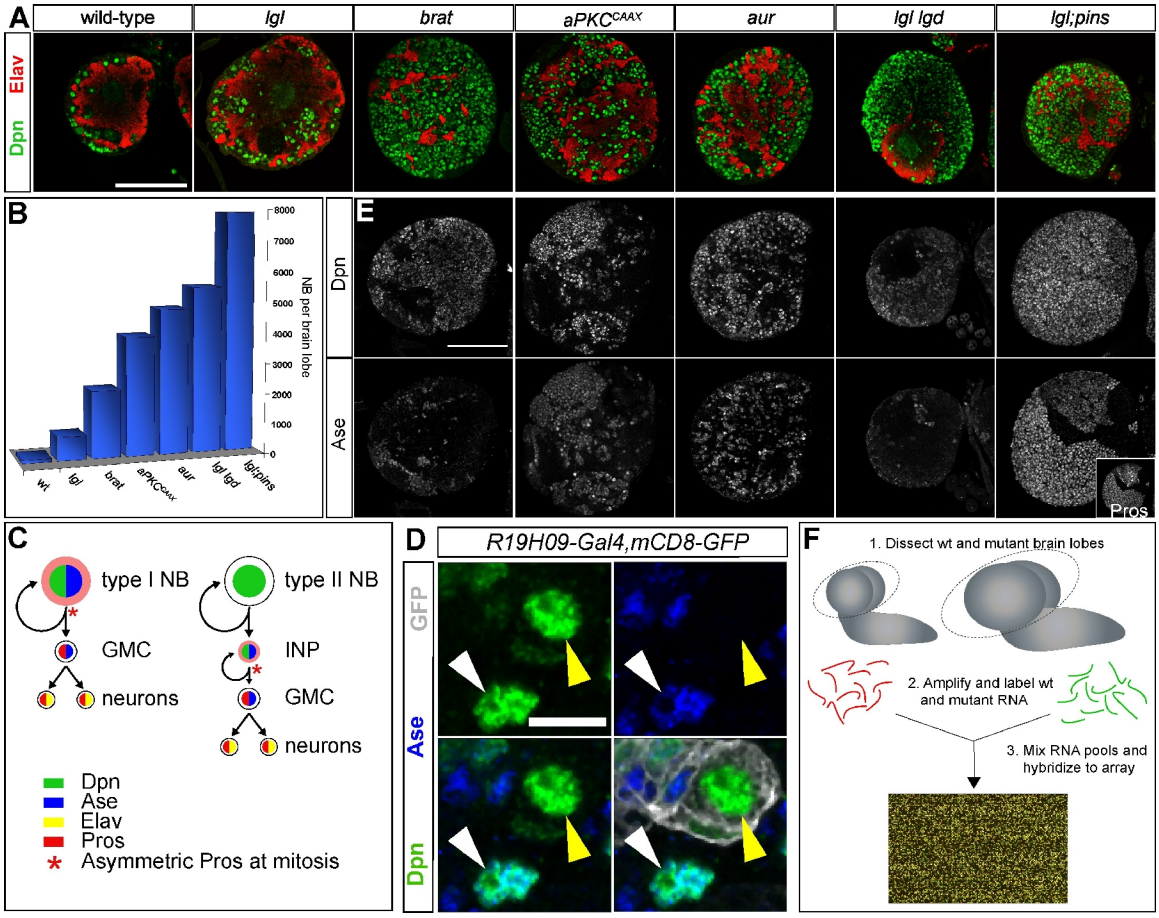

Fig. 8. Using ectopic self-renewal mutants for expression profiling of neuroblasts. Six genotypes were used which are known to cause expansions in the number of neuroblasts in Drosophila larval central brains. (A) Single-slice confocal images of wild-type (120 hours ALH) and mutant (144 hours ALH) brain lobes stained for Dpn (neuroblast marker) and Elav (neuronal marker). (B) The variable level of ectopic neuroblast number per brain lobe of each mutant genotype (n=2 for each genotype). (C) Schematic of wild-type type I and type II neuroblast divisions. Type I neuroblasts have nuclear Ase as well as diffuse cytoplasmic Pros, which is asymmetrically segregated into the GMC upon neuroblast division. Type II neuroblasts lack both Pros and Ase, both of which are expressed in INPs; Pros is then segregated asymmetrically into the GMC upon INP division. GMCs divide to generate Elav^{+} neurons. (D) High magnification image of a Dpn^{+} Ase^{+} type I neuroblast (white arrowhead) and a Dpn^{+} Ase^{-} type II neuroblast (yellow arrowhead) in the dorso-medial region of a wild-type brain. The type II neuroblast can be unambiguously identified based on the presence of GFP driven by *R19H09-Gal4* (Bayraktar et al., 2010). (E) Mutant brains (120 hours ALH) stained with anti-Dpn (to mark all neuroblasts) and anti-Ase (which only marks type I neuroblasts). Inset in the Ase panel of the *lgl pins* brain shows that the Pros staining pattern in the same brain matches very closely to the Ase pattern. (F) Schematic of the methodology used here. Scale bars: 10 µm in (D); 100 µm in (A and E).

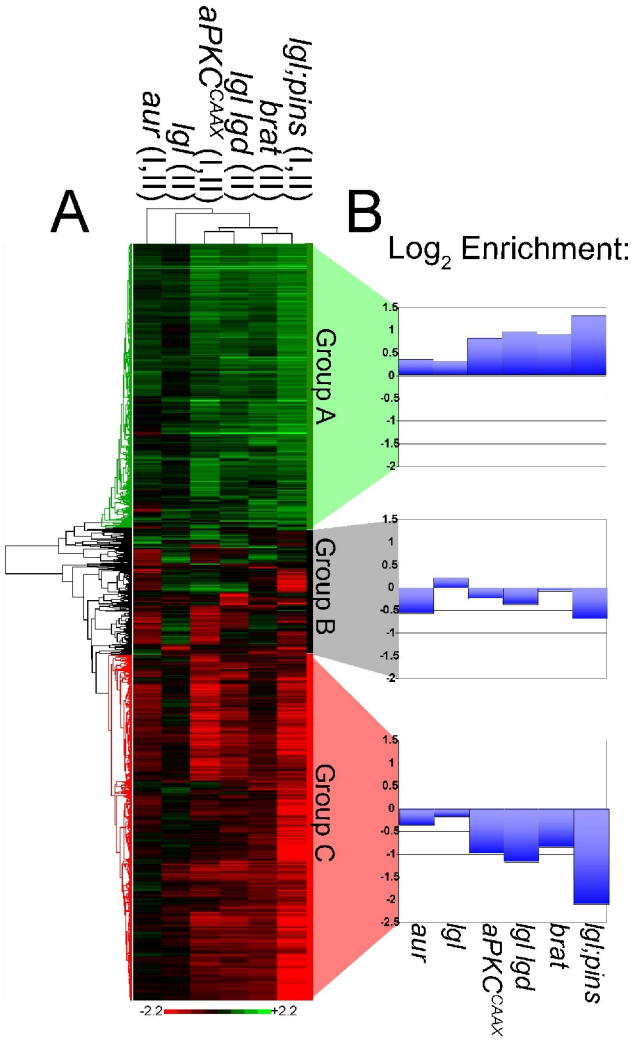

Fig. 9. Results of cluster analysis. (A) Cluster analysis-categorized genes with expression changes in mutant compared to wild-type brains, divided into three groups (A, B, and C).The dendrogram at the top is labeled according to the mutant genotype; Roman numerals indicate the neuroblast subtype(s) enriched in each mutant (I = type I; II = type II). (B) Log_2 expression changes (mutant/wild-type) averaged over all genes in each group.

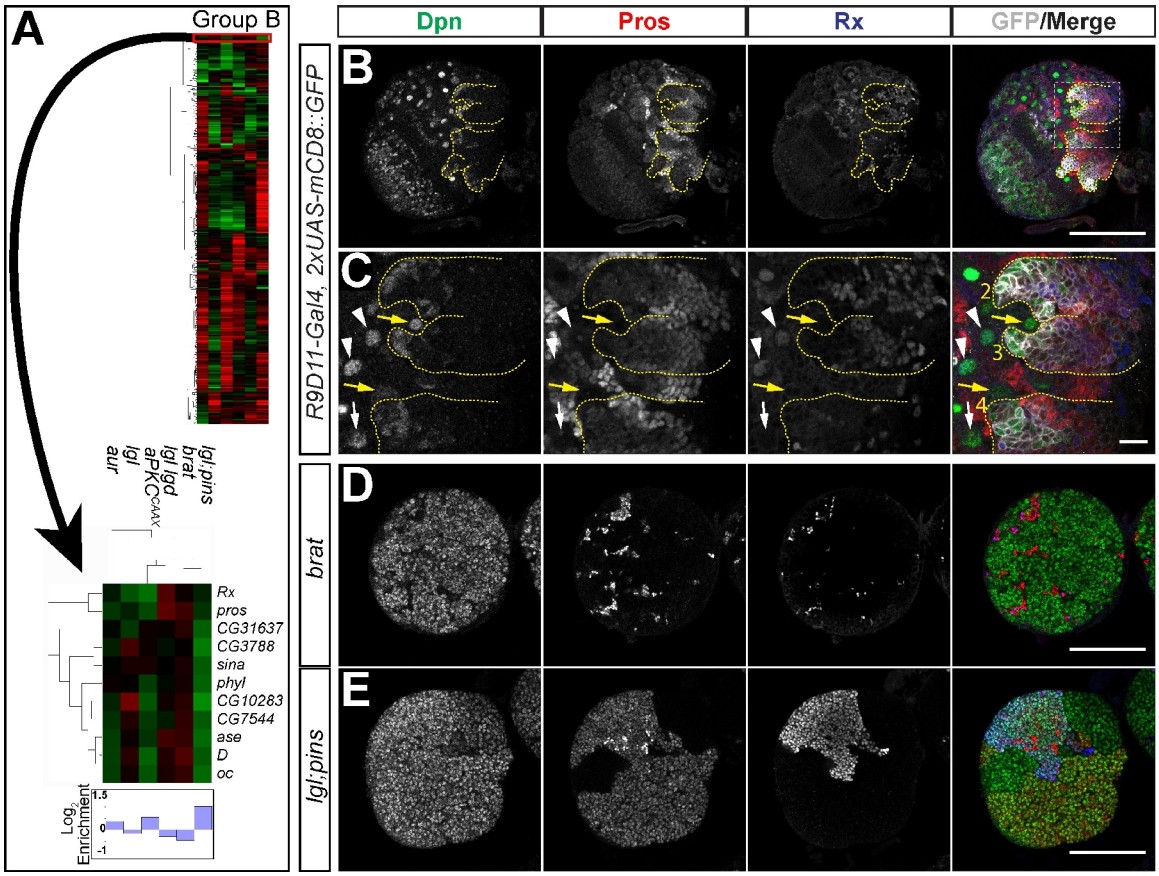

Fig. 10. Differential expression of genes excluded from type II neuroblasts. (A) Position within group B of sub-cluster containing *pros* and *ase* as well as nine other genes with unknown expression patterns. Log_2 enrichment over wild-type is shown, averaged over all genes in the sub-cluster. (B) Confocal image of a wild-type brain lobe at 120 hours ALH. Multiple type I neuroblasts, four type II lineages, and three type II neuroblasts are visible. (C) Enlargement of the region boxed in (B). All neuroblasts are Dpn^+; a subset of type I neuroblasts are Rx^+, while type II neuroblasts are Rx^-. INPs are also Rx^-. Rx is expressed in a subset of neuronal progeny in both type I and type II lineages. (D) *brat* mutant brain lobe (120 hours ALH) contains many Dpn^+ neuroblasts, but these cells do not express Rx. Rx is expressed in a few of the $Pros^+$ cells, all of which in this focal plane are neurons and do not express Dpn. (E) *lgl pins* brain lobe (120 hours ALH) in which Rx is expressed in a subset of ectopic Dpn^+ neuroblasts. Rx expression is limited to cells expressing Pros, which in *lgl pins* also express Ase and are likely derived from expansion of type I neuroblasts. The $Pros^-$ regions (type II-derived) are entirely Rx^-. White arrow: Rx^- type I neuroblast; white arrowheads: Rx^+ type I neuroblasts; yellow arrows: Rx^- type II neuroblasts; mCD8::GFP driven by *R9D11-Gal4* marks a subset of type II lineages, but not the type II neuroblasts themselves (Bayraktar et al., 2010). Shown outlined here with yellow dashed lines are several dorso-medial type II lineages [DM 2, 3, 4, 5, and 6 in (B); DM 2, 3, and 4 in (C)]. Scale bars: 100 µm in (B), (D), and (E); 10 µm in (C).

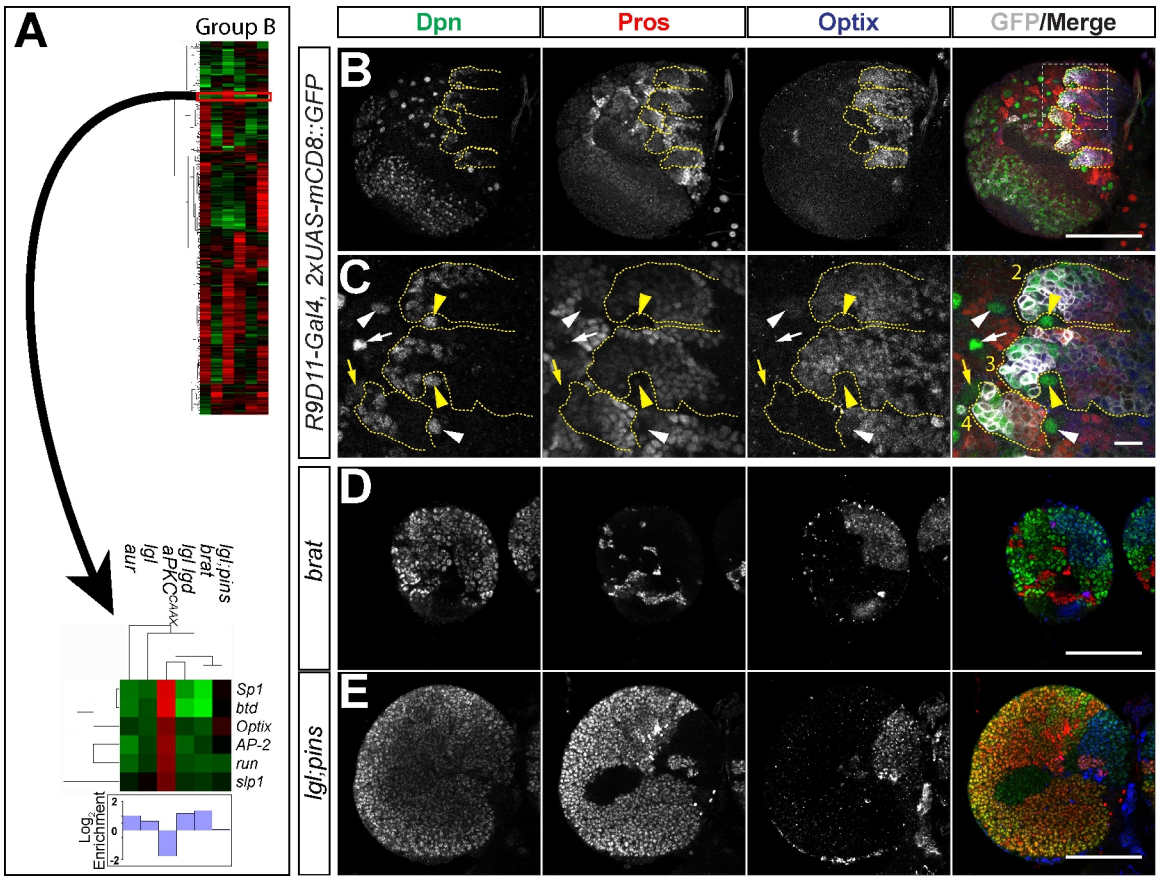

Fig. 11. Identification of a cluster with type II-biased expression. (A) Position in group B of a sub-cluster in which genes are expressed higher in *brat* and *lgl lgd* than in other genotypes. Enrichment shown is averaged over all genes in the sub-cluster. (B) Wild-type brain lobe (120 hours ALH). Visible are multiple type I neuroblasts as well as several type II neuroblasts and their lineages. Most type I neuroblasts are Optix$^-$, while four of the six dorso-medial type II neuroblasts are Optix$^+$ (DM1,2,3,6); Optix is absent from type II neuroblasts DM 4&5 and their INPs, but present in a subset of their progeny. (C) Enlargement of the box in (B) shows both Optix$^+$ and Optix$^-$ type I and type II neuroblasts. Shown are type II lineages DM 2, 3, and 4. Optix is nearly absent from the entire DM4 lineage. (D) *brat* brain (120 hours ALH) shows that Optix is expressed in a dorso-medial region in which nearly all cells are Dpn$^+$ ectopic neuroblasts. (E) *lgl pins* brain (120 hours ALH) exhibits Optix expression primarily in Dpn$^+$ Pros$^-$ regions (type II-derived ectopic neuroblasts). White arrows: Optix$^-$ type I neuroblasts; white arrowheads: Optix$^+$ type I neuroblasts; yellow arrow: Optix$^-$ type II neuroblast; yellow arrowheads: Optix$^+$ type II neuroblasts. GFP driven by *R9D11-Gal4* marks dorso-medial type II lineages, but not the type II neuroblasts themselves. Shown outlined here with yellow dashed lines are several dorso-medial type II lineages [DM 2, 3, 4, 5, and 6 in (B); DM 2, 3, and 4 in (C)]. Scale bars: 100 µm in (B), (D), and (E); 10 µm in (C).

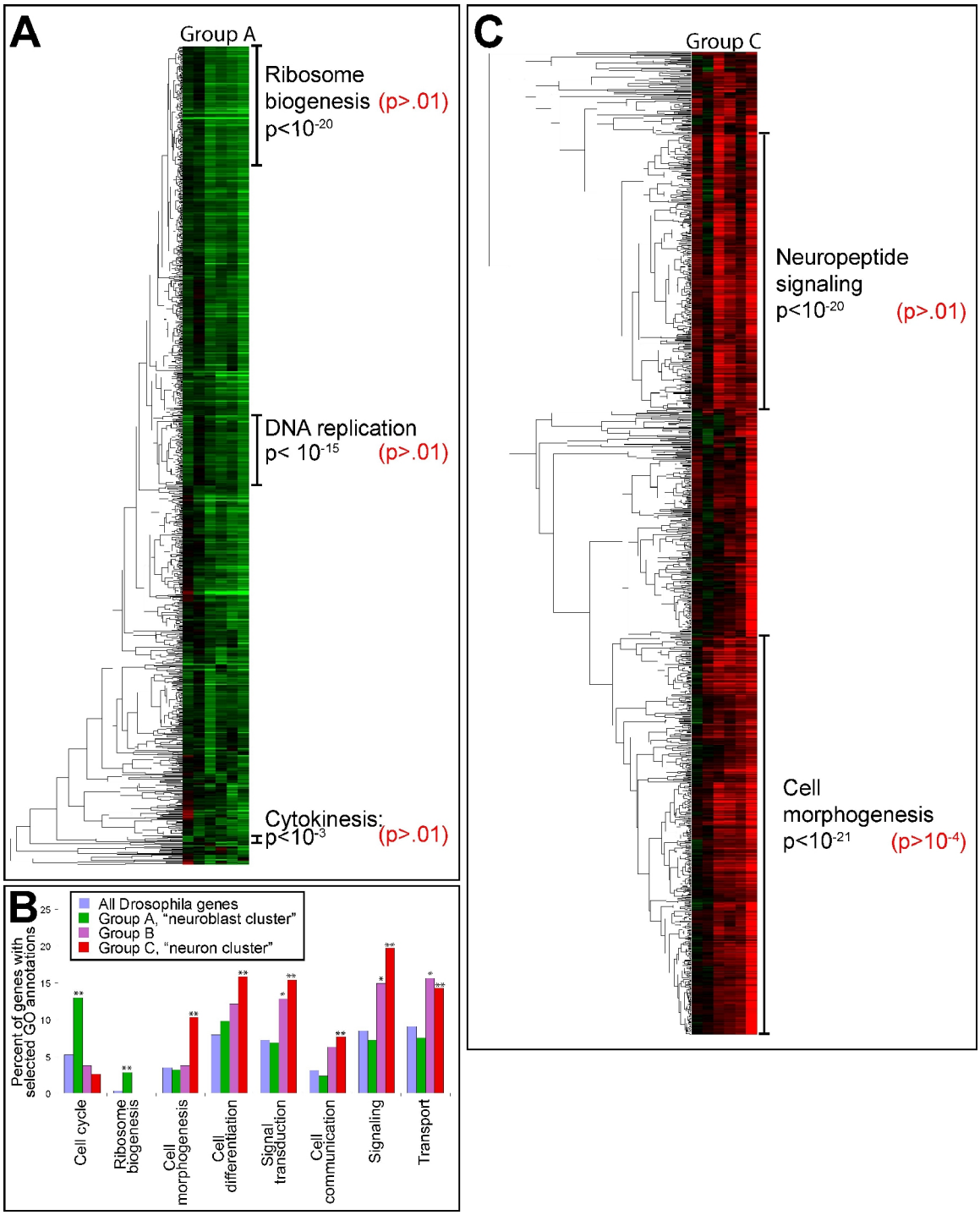

Fig. 12. Gene ontology terms enriched in each group. (A) Group A, the "neuroblast cluster" with three sub-clusters marked in which the indicated GO annotations are significantly enriched compared to all Drosophila genes. Each value in red indicates the enrichment of the GO term in all group A genes *excluding* the adjacent sub-cluster. (B) Chart depicting the percent of all Drosophila genes characterized by select GO annotations as well as percent of genes in each group with those annotations. Asterisks indicate significant enrichment of GO term compared with all Drosophila genes (*:

p<.05; **: p<.001). (C) Group C, the “neuron cluster” with sub-clusters labeled indicating significantly enriched GO terms; each value in red indicates the enrichment of the GO term in all group C genes *except* the adjacent sub-cluster.

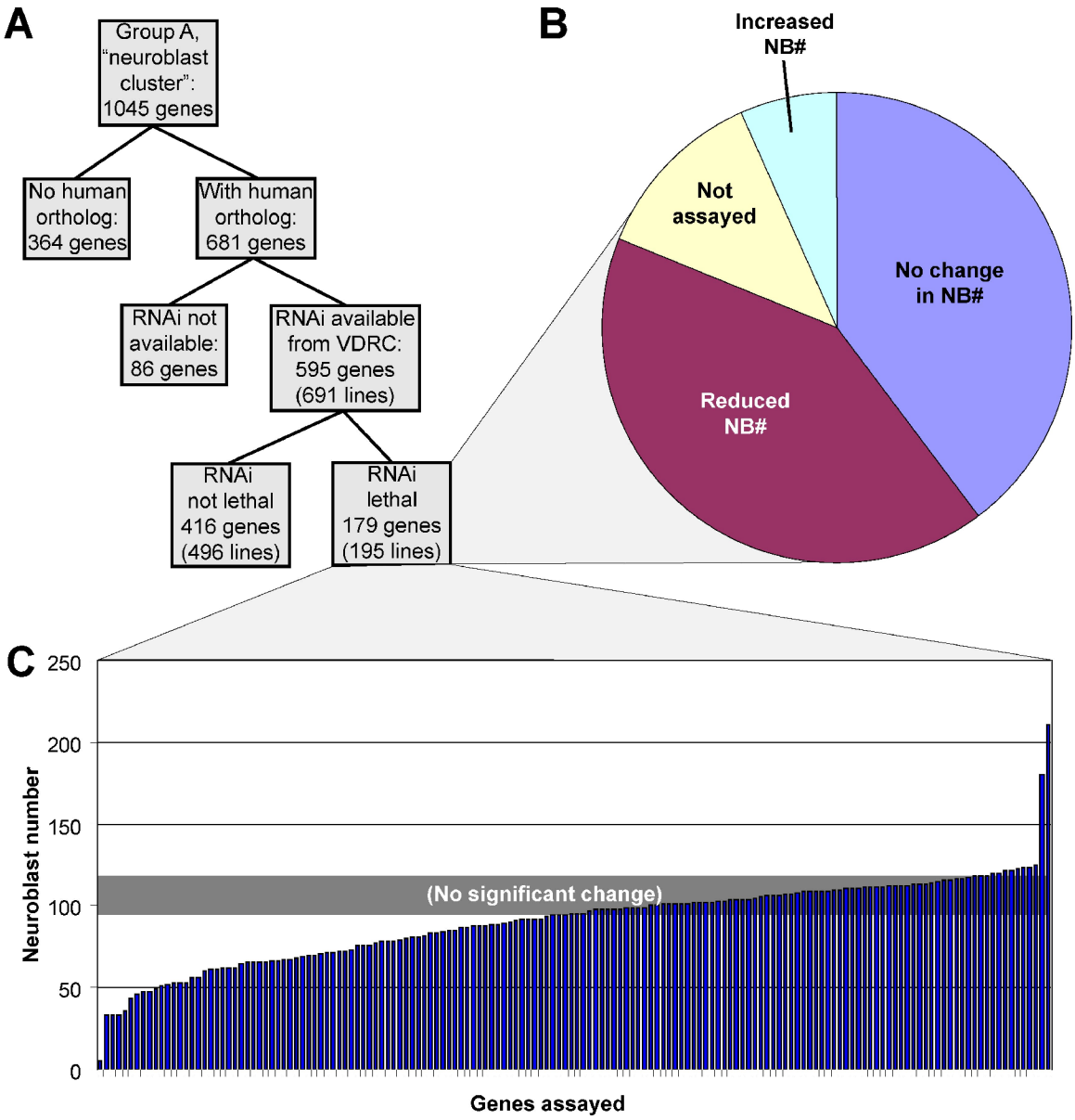

Fig. 13. RNAi screen identifies neuroblast homeostasis genes. (A) Flowchart describing the selection of 691 RNAi lines used in this screen. (B) Neuroblast number gain and loss phenotypes of the 179 genes for which RNAi knock-down caused lethality. (C) Neuroblast numbers per brain lobe of the genes which were assayed for neuroblast number phenotype.

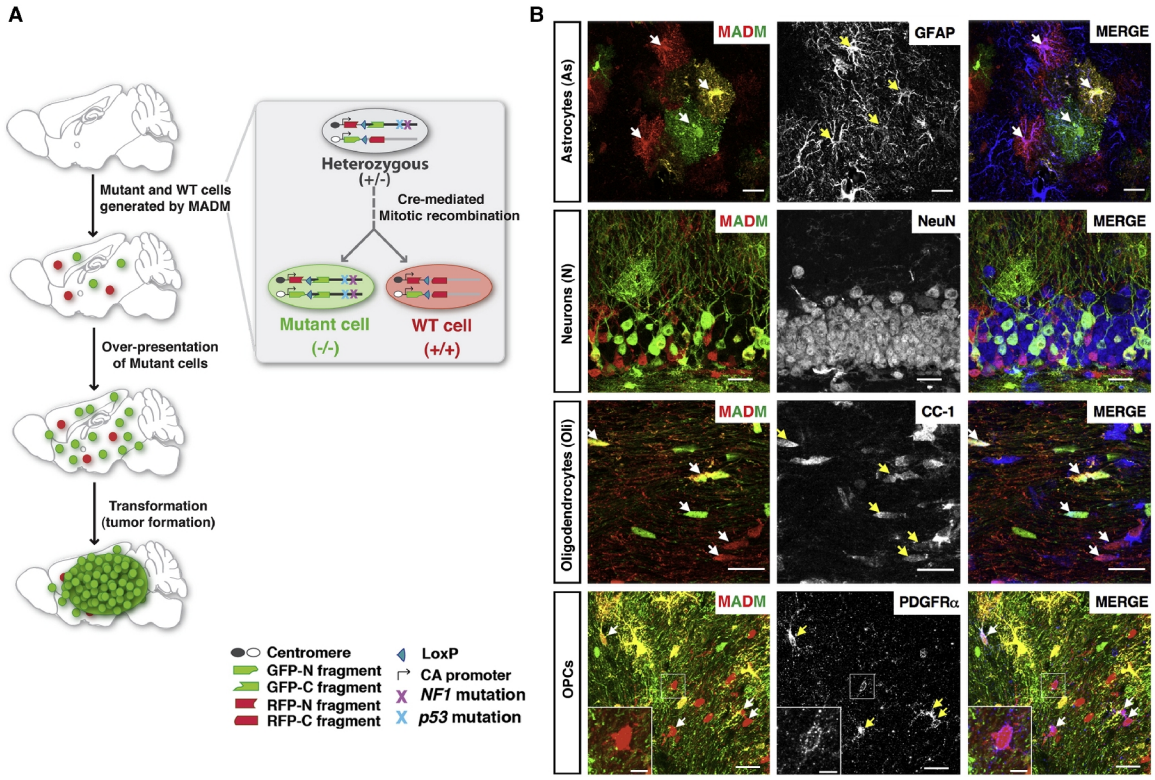

Fig. 14. MADM-based glioma model allows phenotypic analysis at single-cell resolution. (A) Scheme of MADM-based glioma modeling. Inset illustrates how MADM concurrently mutates and labels cells. (B) Representative confocal images show *hGFAP-Cre*-induced MADM labeling of four NSC-derived cell types in a 2-month-old WT-MADM mouse. Arrows point to MADM-labeled cells expressing corresponding markers. Scale bars, 20 μm; inset, 5 μm.

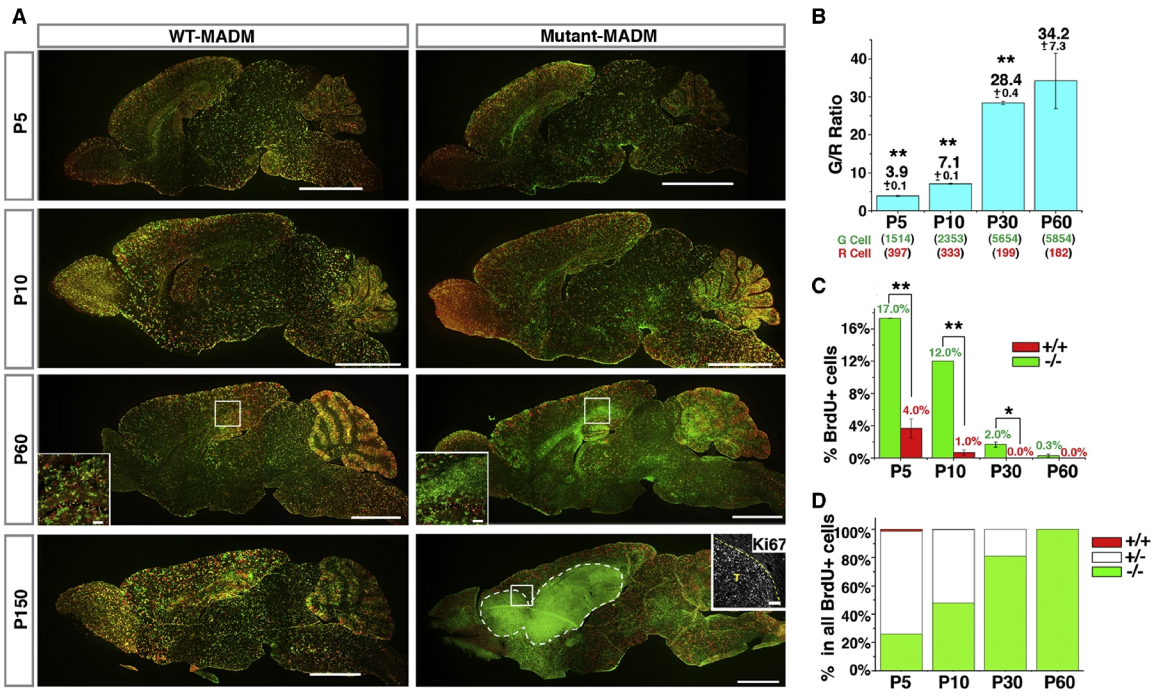

Fig. 15. MADM-mediated sporadic concurrent inactivation of *p53* and *NF1* in embryonic NSCs reveals the entire process of gliomagenesis. (A) Sagittal sections from brains of MADM mice at indicated ages. In WT-MADM mice (left column), both green and red cells are WT. In mutant-MADM mice (right column), green cells are *p53* and *NF1* double null; red cells, WT. Tumor boundary is demarcated with dashed line. Scale bars, 2 mm; insets, 100 μm. Ki67 staining shows that tumor cells are highly proliferative. (B) Systematic quantification of G/R ratios in mutant-MADM brains from P5 to P60. Total cell numbers counted are shown in parentheses. Each number is the sum from three brains. Error bars represent standard error of the mean (SEM). (C) The percentage of BrdU+ cells in WT and mutant cell populations in the brain parenchyma of mutant-MADM mice at indicated ages. Error bars represent SEM. (D) The proportion of BrdU+ cells with genotypes "−/−," "+/−," and "+/+" in the brain parenchyma of mutant-MADM mice at indicated ages. "+/−" includes both double-colored and colorless BrdU+ cells. In (C) and (D), BrdU was administered 1.5 hr prior to sacrifice. Error bars ± SEM; n = 3 mice. $*p < 0.05$, $**p < 0.01$, paired t test.

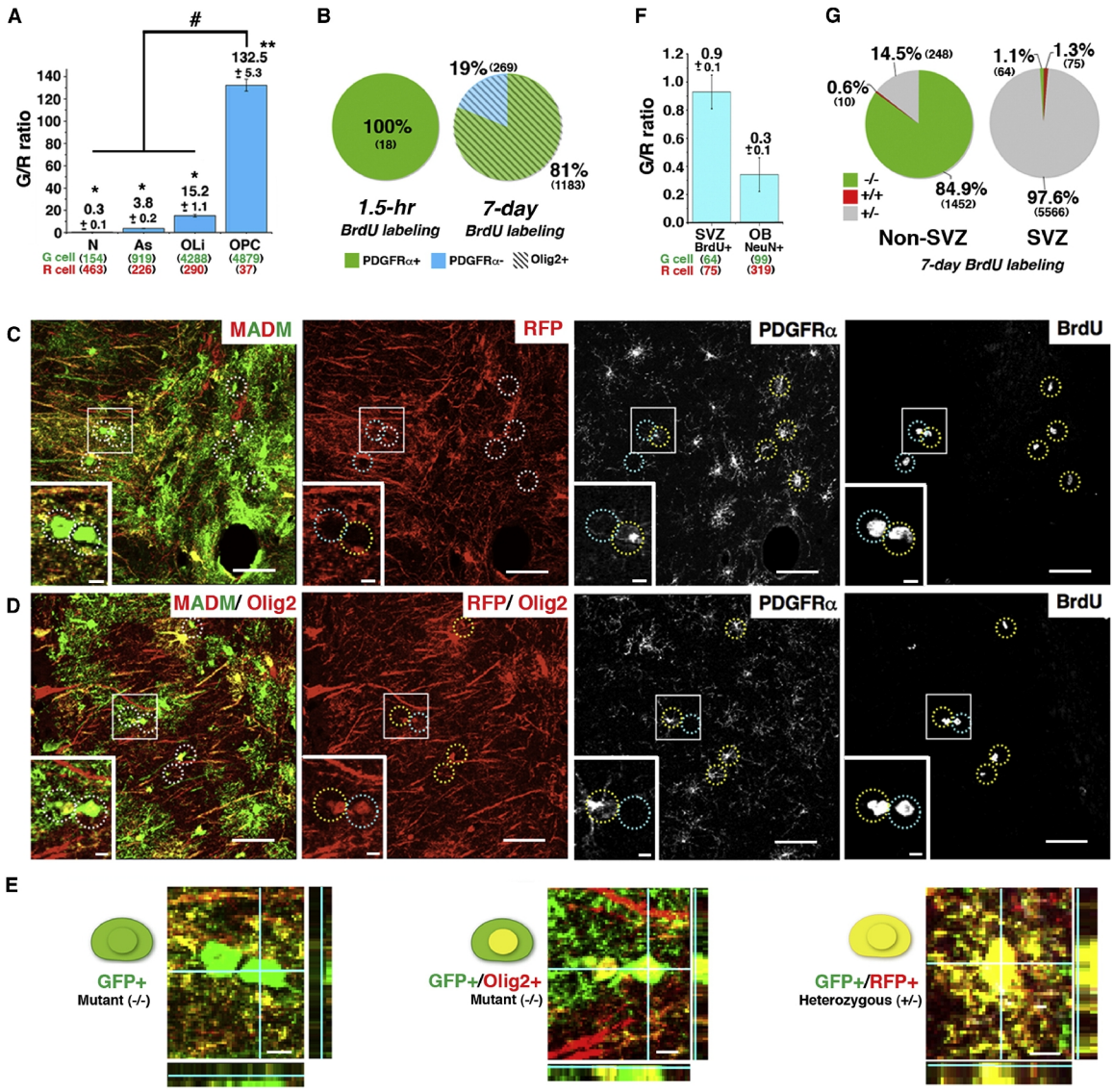

Fig. 16. Analysis at a pretransforming stage of gliomagenesis suggests that OPCs rather than NSCs serve as the cell of origin. (A) Average G/R ratios of each cell type in brain parenchyma of P60 mutant-MADM mice. ∗p < 0.05, ∗∗p < 0.01, paired t test. #p < 0.0001, one-way ANOVA. Error bars represent SEM. (B) (Left chart) With a single BrdU injection (1.5 hr prior to sacrifice) at P60, BrdU-positive mutant (−/−) cells in the brain parenchyma consist entirely of OPCs (PDGFRα+). (Right chart) Upon BrdU administration by drinking water for 1 week, all BrdU-positive mutant (−/−) cells in the brain parenchyma belong to the oligodendrocytic lineage (Olig2+), and the majority of them are OPCs (PDGFRα+). (C–E) "4+1" channel staining shows that all BrdU+ mutant OPCs in a P60 mutant-MADM mouse brain belong to the oligodendrocyte lineage. (C) Without Olig2 staining, all MADM-labeled BrdU+ cells are mutant (green). Notably, some BrdU+ mutant cells are PDGFRα negative (marked with cyan circles). (D) An adjacent section stained with Olig2 together with MADM, BrdU, and PDGFRα shows that all mutant BrdU+ cells have red nuclei, indicating positive staining of Olig2. (E) Representative magnified confocal images show: mutant cells without Olig2 staining

(left), mutant cells with Olig2 staining in red channel (middle), and heterozygous yellow cells (right). The orthogonal z axis is shown on the side of each panel. Scale bars: (C and D) 20 μm; insets, 5 μm; (E) 5 μm. (F) G/R ratios of BrdU+ cells in the SVZ and NeuN+ cells in the olfactory bulb (OB) from P60 mutant-MADM mice. BrdU was given in drinking water for 7 days. (G) Quantification of cells with indicated genotypes among all BrdU+ cells in the non-SVZ brain parenchyma (left chart) or the SVZ (right chart) after one week of BrdU administration. n = 3 mice for all quantification in (A), (B), (F) and (G). Error bars represent ± SEM. Total cell numbers counted in (A), (B), (F), and (G) are shown in parentheses. Each number is the sum from three brains.

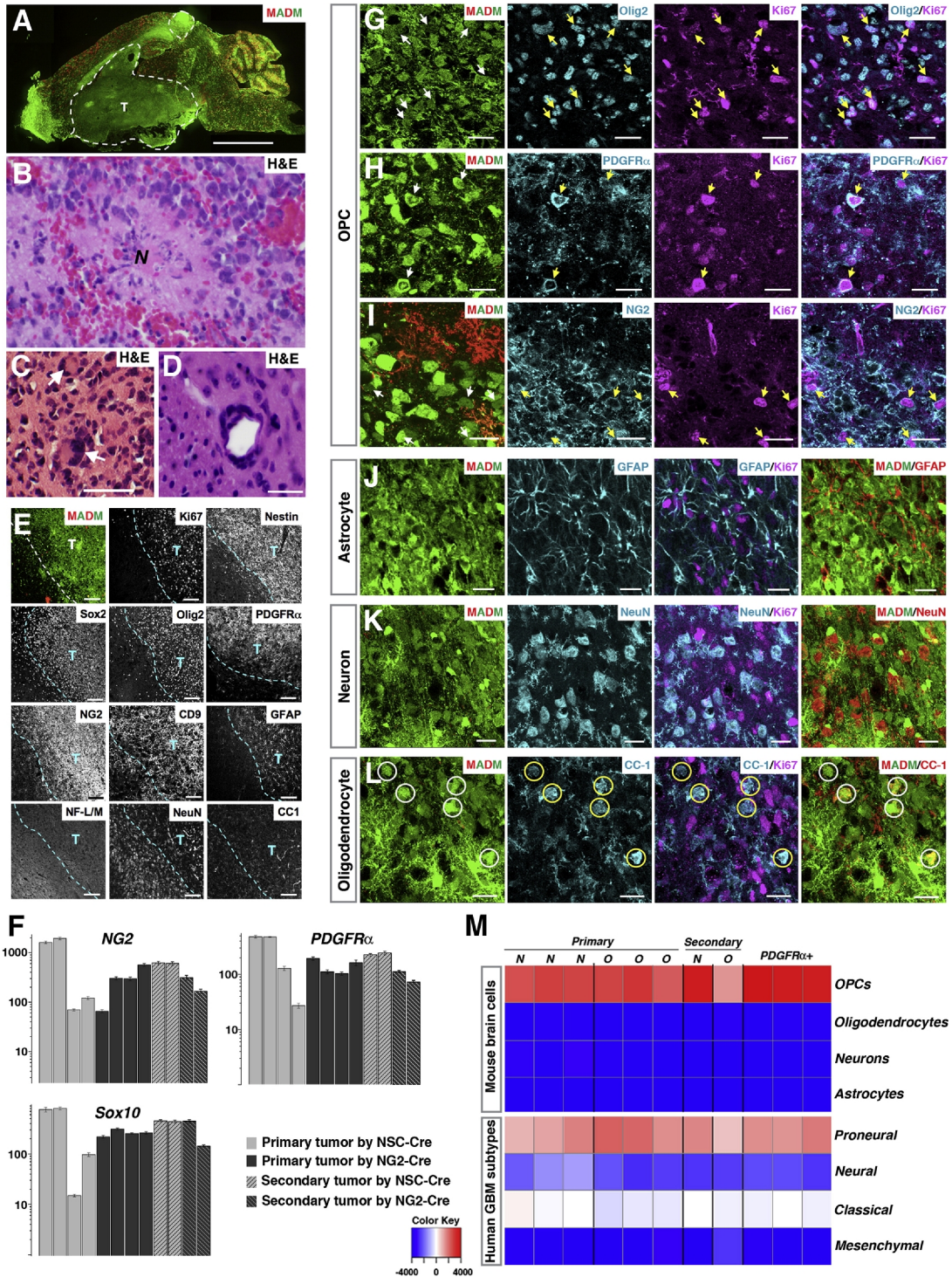

Fig. 17. MADM-generated glioma cells exhibit many OPC features. (A) Representative image of a mutant-MADM brain carrying a GFP+ glioma. (B–D) Adjacent H&E staining of tumor regions shows typical glioma features, including necrotic areas ("*N*" in B), multinucleated giant cells (C), and perivascular satellitosis (D). Scale bars: (A) 2 mm; (C and D) 200μm. Magnification in (B), 400×. (E) Representative low-magnification images show elevated expression of a panel of well-established glioma markers in tumor regions.

All staining was done with adjacent sections from the same tumor. Tumor boundary is demarcated by dashed lines. T, tumor mass. Scale bars, 100 μm except for CD9 staining, for which it is 50 μm. (F) Quantitative RT-PCRs confirm the overexpression of OPC markers in MADM-generated glioma. (G–L) Confocal images at high magnification show that proliferating (Ki67+) green tumor cells express markers for OPCs (G–I, pointed by arrows), but not for other cell types (J–L). The signals of cell type marker staining in the right column of (J–L) were converted to red for better examination of their colocalization with GFP. Some Ki67-negative green cells were CC1+ (circled in L). Scale bars, 20 μm. (M) Transcriptome comparison between tumor samples and four neuroglial cell types (top four rows) and the four subtypes of human GBMs defined by TCGA (bottom four rows) with the single sample Gene Set Enrichment Analysis (ssGSEA) method. *N* and *O* represent tumor samples from mutant-MADM mice induced by *NSC-Cre* (*Nestin-Cre* or *hGFAP-Cre*) and by *NG2-Cre*, respectively. *PDGFRα+* indicates primary tumor cells enriched by anti-PDGFRα immunopanning method. Red to blue indicates significantly similar to dissimilar.

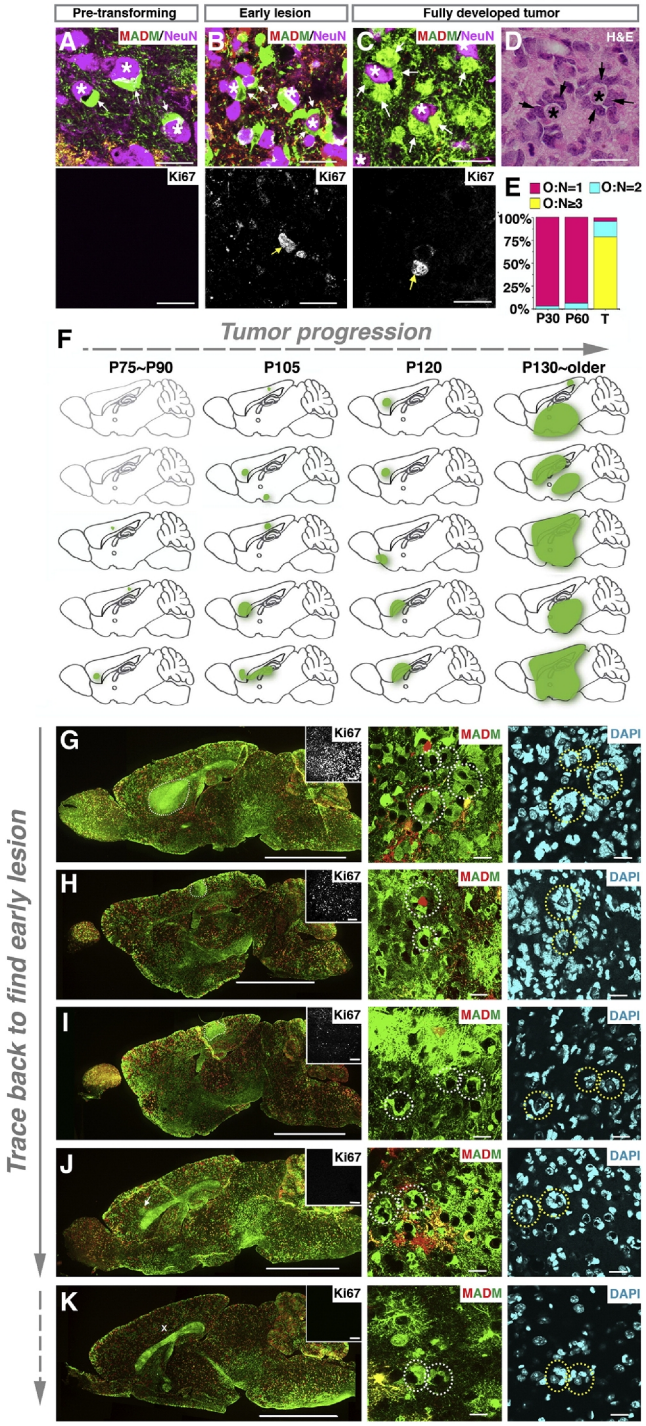

Fig. 18. Spatial analyses of early lesions based on perineuronal cytoarchitecture as a landmark suggest that gliomas initiate at brain regions away from the SVZ. (A–C) Immunofluorescent staining of mutant-MADM brains at distinct tumorigenic stages. Neuronal nuclei were stained with NeuN and are marked as "∗." Arrows point to perineuronal pretransforming OPCs or tumor cells. The proliferating status of perineuronal mutant cells is shown by Ki67 staining (yellow arrows in the bottom row). (D) H&E staining of the adjacent section of (C) shows perineuronal satellitosis. (E)

Proportion of perineuronal structures with distinct mutant OPC-to-neuron ratios (O:N) either in pretransforming MADM mutant brains (P30 and P60, n = 3 brains each) or in tumors (T, n = 4). (F) Schematic summary of lesion sites (green spots), which are defined by O:N ≥ 3 perineuronal structure together with MADM labeling, Ki67 staining, and also pathology in most cases. Brains that are devoid of any detectable lesions are shown in light gray. The analysis is based on a cohort of mutant-MADM mice induced by *hGFAP-Cre*. (G–K) Representative brain images from (F) with gliomas from medium to small sizes. Insets show Ki67 staining of the tumor regions. The glioma identity in these brains was confirmed by pathological criteria except for (K), in which the perineuronal structure of ≥ 3 mutant OPCs suggests that it should be a lesion at its early stage. Scale bars: (A–D) 20 μm; (G–K) left column, 2 mm; middle and right columns, 50 μm.

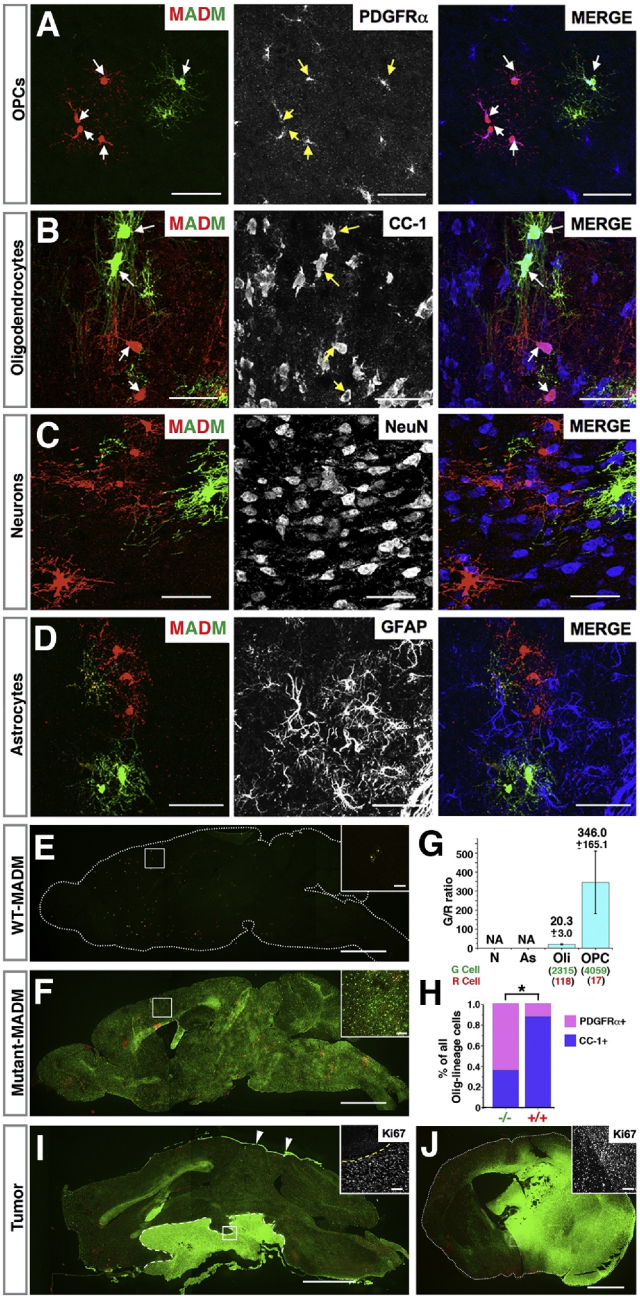

Fig. 19. OPCs can be directly transformed into malignant glioma. (A–D) In the MADM system, NG2-Cre transgene labels OPCs and oligodendrocytes, but not astrocytes or neurons. Arrows point to MADM-labeled cells expressing indicated markers. (E–H) Mutant OPCs overexpand at pretransforming stages. Brain sections from P60 WT-MADM (E) or mutant-MADM (F) mice induced by *NG2-Cre*. (G) G/R ratios within each cell lineage in P60 Mutant-MADM brains induced by *NG2-Cre*. NA, not applicable. Error bars represent ± SEM. Total cell number being counted is shown in parentheses. (H) Percentage of OPCs (PDGFRα+) versus oligodendrocytes (CC1+) within mutant and WT cell populations. n = 3 mice in (G) and (H). ∗p < 0.05, t test. (I and J) Representative gross images of malignant glioma in *NG2-Cre*-induced mutant-MADM mice, either

locating around hypothalamus with invasion into the subarachnoid space (I) or residing within the brain parenchyma (J). Arrows in (I) point to tumor cells spreading along the meninges. Ki67 staining in insets shows that tumor cells are highly proliferative. Scale bars, 2mm; inset, 100 μm.

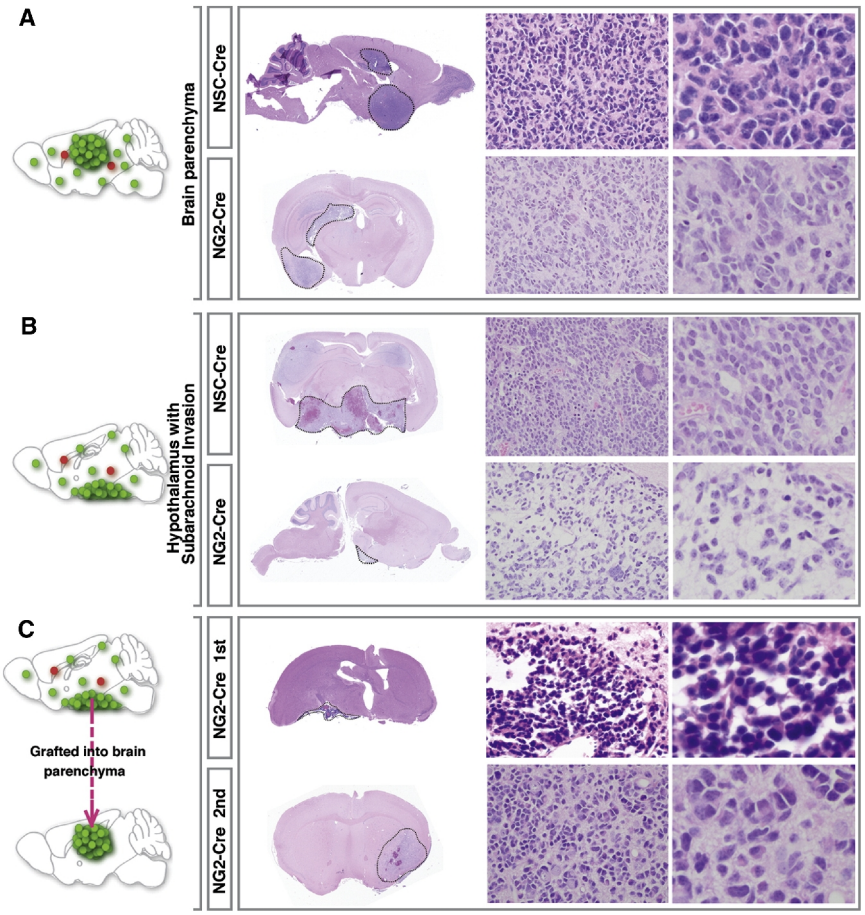

Fig. 20. Comparative pathological analyses of *NSC*- and *NG2-Cre* induced tumors suggest that tumor cell morphology is highly dependent on the location rather than initially mutated cell types. (A and B) Regardless of the Cre lines used, tumors at the same location exhibit indistinguishable pathological features. (C) Transplantation of tumor cells from *NG2-Cre*-induced glioma with subarachnoid invasion (top row) into the brain parenchyma of NOD/SCID mice to generate secondary tumors (bottom row). Pathological features of the primary and the secondary tumors mimic tumor features in (B) and (A), respectively. Tumor boundaries are demarcated by dashed lines. The magnification of images in the middle columns is 400×. Images from the right column are 2.5× digital zoom-in of the corresponding middle-column ones.

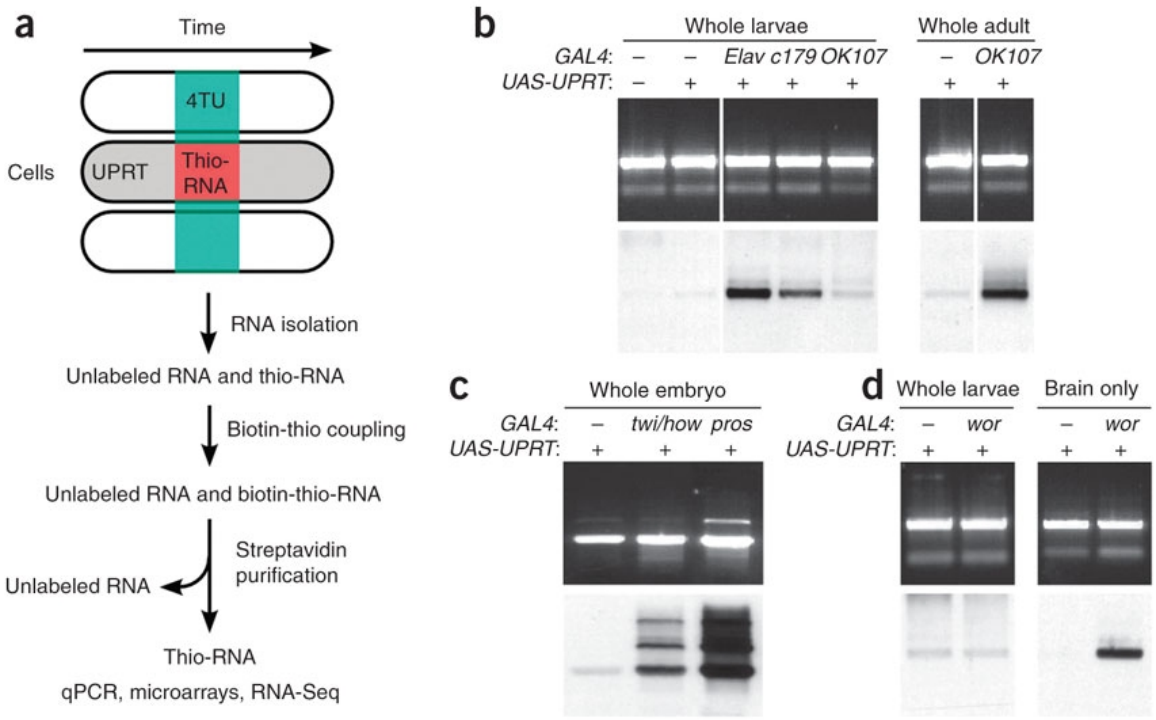

Fig. 21. TU-tagging: overview and cell type-specific labeling. (A) TU-tagging procedure. (B) TU-tagging in larvae and adult flies. RNA from larvae expressing GAL4 in none of the cells (−), all neurons (*elav*), muscle cells (*c179*) or the mushroom body (*OK107*) and either with or without *UAS-UPRT*, and from adults of the indicated genotypes were electrophoresed and stained with ethidium bromide to detect all RNA (top) and streptavidin-horseradish peroxidase to detect thio-RNA (bottom). (C) The 0–16-h embryos of the indicated genotypes were treated with 4-thiouracil for 2 h. RNA was analyzed as indicated above. (D) Comparison of RNA labeling before tissue isolation (whole larvae) and after tissue isolation (brain only). *twi*/*how*, mesoderm/muscle; *pros*, neural; and *wor*, neuroblasts.

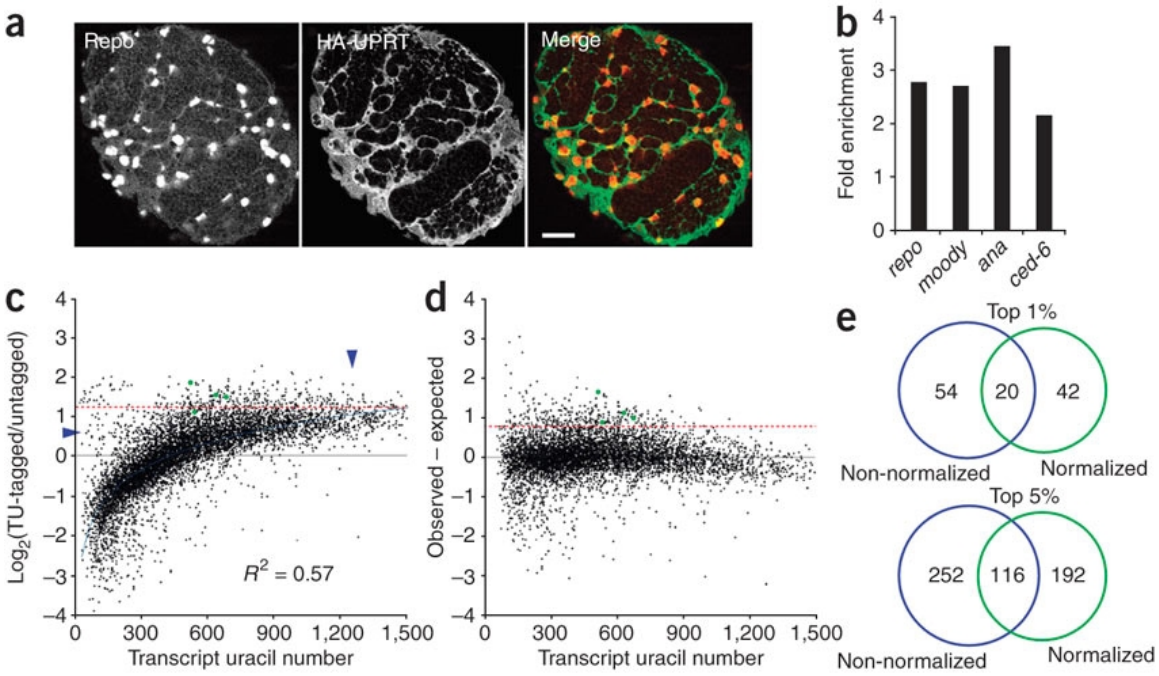

Fig. 22. Cell type-specific RNA isolation and analysis. (A) Single confocal section through a brain lobe from a 96-h after larval hatching (ALH) larvae expressing hemagglutinin-tagged UPRT (HA-UPRT) in all glia (*repo-GAL4 UAS-HA-UPRT*) stained for HA-UPRT (detected with an HA antibody) and glial nuclei (detected with Repo antibody). Scale bar, 20 μm (B) Fold enrichment of the indicated larval glia–specific genes. (C) Average microarray ratios from two glia TU-tagging experiments plotted against the number of uracils in the transcript. Dashed red line indicates cutoff for top 5% enriched genes. Green dots, previously known larval glia–specific genes; vertical arrowhead, possible false positives; and horizontal arrowhead, possible false negatives. (D) TU-tagging microarray ratios after normalization and removal of transcripts with missing untranslated region annotations. (E) Comparison of top 1% and 5% enriched genes before and after normalization for uracil number.

REFERENCES CITED

Akiyama, H., Chaboissier, M.C., Behringer, R.R., Rowitch, D.H., Schedl, A., Epstein, J.A., de Crombrugghe, B., 2004. Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. Proc. Natl. Acad. Sci. U.S.A. 101, 6502-6507.

Albertson, R., Chabu, C., Sheehan, A., Doe, C.Q., 2004. Scribble protein domain mapping reveals a multistep localization mechanism and domains necessary for establishing cortical polarity. J. Cell. Sci. 117, 6061-6070.

Alcantara Llaguno, S., Chen, J., Kwon, C.H., Jackson, E.L., Li, Y., Burns, D.K., Alvarez-Buylla, A., Parada, L.F., 2009. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. Cancer Cell 15, 45-56.

Almeida, M.S., Bray, S.J., 2005. Regulation of post-embryonic neuroblasts by Drosophila Grainyhead. Mech. Dev. 122, 1282-1293.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., et al., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711-1714.

Arango, N.A., Lovell-Badge, R., Behringer, R.R., 1999. Targeted mutagenesis of the endogenous mouse Mis gene promoter: in vivo definition of genetic pathways of vertebrate sexual development. Cell 99, 409-419.

Ashraf, S.I., Ganguly, A., Roote, J., Ip, Y.T., 2004. Worniu, a Snail family zinc-finger protein, is required for brain development in Drosophila. Dev. Dyn. 231, 379-386.

Assanah, M., Lochhead, R., Ogden, A., Bruce, J., Goldman, J., Canoll, P., 2006. Glial progenitors in adult white matter are driven to form malignant gliomas by platelet-derived growth factor-expressing retroviruses. J. Neurosci. 26, 6781-6790.

Atwood, S.X., Prehoda, K.E., 2009. aPKC phosphorylates Miranda to polarize fate determinants during neuroblast asymmetric cell division. Curr. Biol. 19, 723-729.

Awasaki, T., Tatsumi, R., Takahashi, K., Arai, K., Nakanishi, Y., Ueda, R., Ito, K., 2006. Essential role of the apoptotic cell engulfment genes draper and ced-6 in programmed axon pruning during Drosophila metamorphosis. Neuron 50, 855-867.

Bachoo, R.M., Maher, E.A., Ligon, K.L., Sharpless, N.E., Chan, S.S., You, M.J., Tang, Y., DeFrances, J., Stover, E., Weissleder, R., et al., 2002. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. Cancer Cell 1, 269-277.

Bainton, R.J., Tsai, L.T.Y., Schwabe, T., DeSalvo, M., Gaul, U., Heberlein, U., 2005. moody encodes two GPCRs that regulate cocaine behaviors and blood-brain barrier permeability in Drosophila. Cell 123, 145-156.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., et al., 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 462, 108-112.

Barrionuevo, F., Naumann, A., Bagheri-Fam, S., Speth, V., Taketo, M.M., Scherer, G., Neubüser, A., 2008. Sox9 is required for invagination of the otic placode in mice. Dev. Biol. 317, 213-224.

Barrionuevo, F., Taketo, M.M., Scherer, G., Kispert, A., 2006. Sox9 is required for notochord maintenance in mice. Dev. Biol. 295, 128-140.

Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F., et al., 2007. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J. Cell Biol. 178, 635-648.

Bayraktar, O.A., Boone, J.Q., Drummond, M.L., Doe, C.Q., 2010. Drosophila type II neuroblast lineages keep Prospero levels low to generate large clones that contribute to the adult brain central complex. Neural Dev. 5, 26.

Bell, D.M., Leung, K.K., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P., Cheah, K.S., 1997. SOX9 directly regulates the type-II collagen gene. Nat. Genet. 16, 174-178.

Bello, B.C., Izergina, N., Caussinus, E., Reichert, H., 2008. Amplification of neural stem cell proliferation by intermediate progenitor cells in Drosophila brain development. Neural Dev. 3, 5.

Bennett, M.R., Rizvi, T.A., Karyala, S., McKinnon, R.D., Ratner, N., 2003. Aberrant growth and differentiation of oligodendrocyte progenitors in neurofibromatosis type 1 mutants. J. Neurosci. 23, 7207-7217.

Bernardos, R.L., Lentz, S.I., Wolfe, M.S., and Raymond, P.A., 2005. Notch-Delta signaling is required for spatial patterning and Müller glia differentiation in the zebrafish retina. Dev. Biol. 278, 381-395.

Betschinger, J., Mechtler, K., Knoblich, J.A., 2006. Asymmetric segregation of the tumor suppressor brat regulates self-renewal in Drosophila neural stem cells. Cell 124, 1241-1253.

Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., de Crombrugghe, B., 1999. Sox9 is required for cartilage formation. Nat. Genet. 22, 85-89.

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., de Crombrugghe, B., 2001. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc. Natl. Acad. Sci. U.S.A. 98, 6698-6703.

Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L.Y., Jan, Y.N., 1992. deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. Genes Dev. 6, 2137-2151.

Blackshaw, S., Fraioli, R.E., Furukawa, T., Cepko, C.L., 2001. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. Cell 107, 579-589.

Boone, J.Q., Doe, C.Q., 2008. Identification of Drosophila type II neuroblast lineages containing transit amplifying ganglion mother cells. Dev. Neurobiol. 68, 1185-1195.

Bowles, J., Schepers, G., Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev. Biol. 227, 239-255.

Bowman, S.K., Rolland, V., Betschinger, J., Kinsey, K.A., Emery, G., Knoblich, J.A., 2008. The tumor suppressors Brat and Numb regulate transit-amplifying neuroblast lineages in Drosophila. Dev. Cell 14, 535-546.

Brand, M., Jarman, A.P., Jan, L.Y., Jan, Y.N., 1993. asense is a Drosophila neural precursor gene and is capable of initiating sense organ formation. Development 119, 1-17.

Bridgewater, L.C., Lefebvre, V., de Crombrugghe, B., 1998. Chondrocyte-specific enhancer elements in the Col11a2 gene resemble the Col2a1 tissue-specific enhancer. J. Biol. Chem. 273, 14998-15006.

Broadus, J., Fuerstenberg, S., Doe, C.Q., 1998. Staufen-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. Nature 391, 792-795.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al., 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. 28, 264-278.

Caldwell, M.C., Datta, S., 1998. Expression of cyclin E or DP/E2F rescues the G1 arrest of trol mutant neuroblasts in the Drosophila larval central nervous system. Mech. Dev. 79, 121-130.

Catchen, J.M., Conery, J.S., Postlethwait, J.H., 2008. Inferring ancestral gene order. Methods Mol. Biol. 452, 365-383.

Chaboissier, M.C., Kobayashi, A., Vidal, V.I.P., Lützkendorf, S., van de Kant, H.J.G., Wegner, M., de Rooij, D.G., Behringer, R.R., Schedl, A., 2004. Functional analysis of Sox8 and Sox9 during sex determination in the mouse. Development 131, 1891-1901.

Chabu, C. Doe, C.Q., 2008. Dap160/intersectin binds and activates aPKC to regulate cell polarity and cell cycle progression. Development 135, 2739-2746.

Chae, J.H., Stein, G.H., Lee, J.E., 2004. NeuroD: the predicted and the surprising. Mol. Cells 18, 271-288.

Cheung, M., Chaboissier, M.C., Mynett, A., Hirst, E., Schedl, A., Briscoe, J., 2005. The transcriptional control of trunk neural crest induction, survival, and delamination. Dev. Cell 8, 179-192.

Chia, W., Somers, W.G., Wang, H., 2008. Drosophila neuroblast asymmetric divisions: cell cycle regulators, asymmetric protein localization, and tumorigenesis. J. Cell Biol. 180, 267-272.

Chiang, E.F., Pai, C.I., Wyatt, M., Yan, Y.L., Postlethwait, J., Chung, B., 2001. Two sox9 genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Dev. Biol. 231, 149-163.

Choksi, S.P., Southall, T.D., Bossing, T., Edoff, K., de Wit, E., Fischer, B.E., van Steensel, B., Micklem, G., Brand, A.H., 2006. Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells. Dev. Cell 11, 775-789.

Chow, R.L., Volgyi, B., Szilard, R.K., Ng, D., McKerlie, C., Bloomfield, S.A., Birch, D.G., McInnes, R.R., 2004. Control of late off-center cone bipolar cell differentiation and visual signaling by the homeobox gene Vsx1. Proc. Natl. Acad. Sci. U.S.A. 101, 1754-1759.

Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., Boothroyd, J.C., 2005. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat. Biotechnol. 23, 232-237.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., Largaespada, D.A., 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 436, 272-276.

Cresko, W.A., Yan, Y.L., Baltrus, D.A., Amores, A., Singer, A., Rodríguez-Marí, A., Postlethwait, J.H., 2003. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev. Dyn. 228, 480-489.

Davis, R.J., Tavsanli, B.C., Dittrich, C., Walldorf, U., Mardon, G., 2003. Drosophila retinal homeobox (drx) is not required for establishment of the visual system, but is required for brain and clypeus development. Dev. Biol. 259, 272-287.

Dawson, M.R.L., Polito, A., Levine, J.M., Reynolds, R., 2003. NG2-expressing glial progenitor cells: an abundant and widespread population of cycling cells in the adult rat CNS. Mol. Cell. Neurosci. 24, 476-488.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al., 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448, 151-156.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., Xu, T., 2005. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122, 473-483.

Doe, C.Q., 2008. Neural stem cells: balancing self-renewal with differentiation. Development 135, 1575-1587.

Doetsch, F., Caillé, I., Lim, D.A., García-Verdugo, J.M., Alvarez-Buylla, A., 1999. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. Cell 97, 703-716.

Domínguez, M., Campuzano, S., 1993. asense, a member of the Drosophila achaete-scute complex, is a proneural and neural differentiation gene. EMBO J. 12, 2049-2060.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al., 2008. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 135, 749-762.

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al., 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA 14, 1959-1972.

Ebens, A.J., Garren, H., Cheyette, B.N., Zipursky, S.L., 1993. The Drosophila anachronism locus: a glycoprotein secreted by glia inhibits neuroblast proliferation. Cell 74, 15-27.

Eggert, T., Hauck, B., Hildebrandt, N., Gehring, W.J., Walldorf, U., 1998. Isolation of a Drosophila homolog of the vertebrate homeobox gene Rx and its possible role in brain and eye development. Proc. Natl. Acad. Sci. U.S.A. 95, 2343-2348.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95, 14863-14868.

Fadool, J.M., Dowling, J. E., 2008. Zebrafish: a model system for the study of eye genetics. Prog. Retin. Eye Res. 27, 89-110.

Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., et al., 2008. Ensembl 2008. Nucleic Acids Res. 36, D707-714.

Force, A., Lynch, M., Pickett, F.B., Amores, A, Yan, Y.L., Postlethwait, J.H., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531-1545.

Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanović, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N., 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 372, 525-530.

Furukawa, T., Morrow, E.M., Li, T., Davis, F.C., Cepko, C.L., 1999. Retinopathy and attenuated circadian entrainment in Crx-deficient mice. Nat. Genet. 23, 466-470.

Furukawa, T., Morrow, E.M., Cepko, C.L., 1997. Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. Cell 91, 531-541.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543-548.

Ge, H., Liu, Z., Church, G.M., Vidal, M., 2001. Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat. Genet. 29, 482-486.

Geha, S., Pallud, J., Junier, M.P., Devaux, B., Leonard, N., Chassoux, F., Chneiweiss, H., Daumas-Duport, C., Varlet, P., 2010. NG2+/Olig2+ cells are the major cycle-related cell population of the adult human normal brain. Brain Pathol. 20, 399-411.

Genove, G., DeMarco, U., Xu, H., Goins, W.F., Ahrens, E.T., 2005. A new transgene reporter for in vivo magnetic resonance imaging. Nat. Med. 11, 450-454.

Genzer, M.A., Bridgewater, L.C., 2007. A Col9a1 enhancer element activated by two interdependent SOX9 dimers. Nucleic Acids Res. 35, 1178-1186.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al., 2008. A translational profiling approach for the molecular characterization of CNS cell types. Cell 135, 738-748.

Hippenmeyer, S., Youn, Y.H., Moon, H.M., Miyamichi, K., Zong, H., Wynshaw-Boris, A., Luo, L., 2010. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. Neuron 68, 695-709.

Hirata, J., Nakagoshi, H., Nabeshima, Y., Matsuzaki, F., 1995. Asymmetric segregation of the homeodomain protein Prospero during Drosophila development. Nature 377, 627-630.

Hoheisel, J.D., 2006. Microarray technology: beyond transcript profiling and genotype analysis. Nat. Rev. Genet. 7, 200-210.

Hosoya, T., Takizawa, K., Nitta, K., Hotta, Y., 1995. glial cells missing: a binary switch between neuronal and glial determination in Drosophila. Cell 82, 1025-1036.

Houston, C.S., Opitz, J.M., Spranger, J.W., Macpherson, R.I., Reed, M.H., Gilbert, E.F., Herrmann, J., Schinzel, A., 1983. The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971. Am. J. Med. Genet. 15, 3-28.

Ihanamäki, T., Säämänen, A.M., Suominen, J., Pelliniemi, L.J., Harley, V., Vuorio, E., Salminen, H., 2002. Expression of Sox9 and type IIA procollagen during ocular development and aging in transgenic Del1 mice with a mutation in the type II collagen gene. Eur. J. Ophthalmol. 12, 450-458.

Izergina, N., Balmer, J., Bello, B., Reichert, H., 2009. Postembryonic development of transit amplifying neuroblast lineages in the Drosophila brain. Neural. Dev. 4, 44.

Jacks, T, Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T., Weinberg, R.A., 1994. Tumor spectrum analysis in p53-mutant mice. Curr. Biol. 4, 1-7.

Jaekel, R., Klein, T., 2006. The Drosophila Notch inhibitor and tumor suppressor gene lethal (2) giant discs encodes a conserved regulator of endosomal trafficking. Dev. Cell 11, 655-669.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., et al., 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431, 946-957.

Jansen, R., Greenbaum, D., Gerstein, M., 2002. Relating whole-genome expression data with protein-protein interactions. Genome Res. 12, 37-46.

Jensen, A.M., Walker, C., Westerfield, M., 2001. mosaic eyes: a zebrafish gene required in pigmented epithelium for apical localization of retinal cell division and lamination. Development 128, 95-105.

Kenyon, K.L., Li, D.J., Clouser, C., Tran, S., Pignoni, F., 2005. Fly SIX-type homeodomain proteins Sine oculis and Optix partner with different cofactors during eye development. Dev. Dyn. 234, 497-504.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev. Dyn. 203, 253-310.

Kist, R., Schrewe, H., Balling, R., Scherer, G., 2002. Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. Genesis 32, 121-123.

Knoblich, J.A., Jan, L.Y., Jan, Y.N., 1995. Asymmetric segregation of Numb and Prospero during cell division. Nature 377, 624-627.

Knoblich, J.A., 2008. Mechanisms of asymmetric stem cell division. Cell 132, 583-597.

Knudson, A.G., Jr., 1971. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. U.S.A. 68, 820-823.

Komitova, M., Zhu, X., Serwanski, D.R., Nishiyama, A., 2009. NG2 cells are distinct from neurogenic cells in the postnatal mouse subventricular zone. J. Comp. Neurol. 512, 702-716.

Kondo, T., Raff, M., 2000. Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. Science 289, 1754-1757.

Korzh, V., Sleptsova, I., Liao, J., He, J., Gong, Z., 1998. Expression of zebrafish bHLH genes ngn1 and nrd defines distinct stages of neural differentiation. Dev. Dyn. 213, 92-104.

Lee, C.Y., Andersen, R.O., Cabernard, C., Manning, L., Tran, K.D., Lanskey, M.J., Bashirullah, A., Doe, C.Q., 2006. Drosophila Aurora-A kinase inhibits neuroblast self-renewal by regulating aPKC/Numb cortical polarity and spindle orientation. Genes Dev. 20, 3464-3474.

Lee, C.Y., Robinson, K.J., Doe, C.Q., 2006. Lgl, Pins and aPKC regulate neuroblast self-renewal versus differentiation. Nature 439, 594-598.

Lee, C.Y., Wilkinson, B.D., Siegrist, S.E., Wharton, R.P., Doe, C.Q., 2006. Brat is a Miranda cargo protein that promotes neuronal differentiation and inhibits neuroblast self-renewal. Dev. Cell 10, 441-449.

Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., Weintraub, H., 1995. Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. Science 268, 836-844.

Lee, J.S., Padmanabhan, A., Shin, J., Zhu, S., Guo, F., Kanki, J.P., Epstein, J.A., Look, A.T., 2010. Oligodendrocyte progenitor cell numbers and migration are regulated by the zebrafish orthologs of the NF1 tumor suppressor gene. Hum. Mol. Genet. 19, 4643-4653.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., de Crombrugghe, B., 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol. Cell. Biol. 17, 2336-2346.

Levine, E.M., Passini, M., Hitchcock, P.F., Glasgow, E., Schechter, N., 1997. Vsx-1 and Vsx-2: two Chx10-like homeobox genes expressed in overlapping domains in the adult goldfish retina. J. Comp. Neurol. 387, 439-448.

Li, L., Vaessin, H., 2000. Pan-neural Prospero terminates cell proliferation during Drosophila neurogenesis. Genes Dev. 14, 147-151.

Ligon, K.L., Huillard, E., Mehta, S., Kesari, S., Liu, H., Alberta, J.A., Bachoo, R.M., Kane, M., Louis, D.N., Depinho, R.A., et al., 2007. Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. Neuron 53, 503-517.

Lincoln, J., Kist, R., Scherer, G., Yutzey, K.E., 2007. Sox9 is required for precursor cell expansion and extracellular matrix organization during mouse heart valve development. Dev. Biol. 305, 120-132.

Lindberg, N., Kastemar, M., Olofsson, T., Smits, A., Uhrbom, L., 2009. Oligodendrocyte progenitor cells can act as cell of origin for experimental glioma. Oncogene 28, 2266-2275.

Liu, Y., Li, H., Tanaka, K., Tsumaki, N., Yamada, Y., 2000. Identification of an enhancer sequence within the first intron required for cartilage-specific transcription of the alpha2(XI) collagen gene. J. Biol. Chem. 275, 12712-12718.

Liu, Y., Shen, Y., Rest, J.S., Raymond, P.A., Zack, D.J., 2001. Isolation and characterization of a zebrafish homologue of the cone rod homeobox gene. Invest. Ophthalmol. Vis. Sci. 42, 481-487.

Livesey, F.J., Furukawa, T., Steffen, M.A., Church, G.M., Cepko, C.L. (2000). Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. Curr. Biol. 10, 301-310.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., 2007. WHO Classification of Tumours of the Central Nervous System 4th ed. WHO Press, Geneva, Switzerland.

Malicki, J., 2000. Harnessing the power of forward genetics--analysis of neuronal diversity and patterning in the zebrafish retina. Trends Neurosci. 23, 531-541.

Mansour, S., Offiah, A.C., McDowall, S., Sim, P., Tolmie, J., Hall, C., 2002. The phenotype of survivors of campomelic dysplasia. J. Med. Genet. 39, 597-602.

Marshall, O.J., Harley, V.R., 2000. Molecular mechanisms of SOX9 action. Mol. Genet. Metab. 71, 455-462.

Mavropoulos, A., Devos, N., Biemar, F., Zecchin, E., Argenton, F., Edlund, H., Motte, P., Martial, J.A., Peers, B., 2005. sox4b is a key player of pancreatic alpha cell differentiation in zebrafish. Dev. Biol. 285, 211-223.

McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061-1068.

Merkle, F.T., Alvarez-Buylla, A., 2006. Neural stem cells in mammalian development. Curr. Opin. Cell Biol. 18, 704-709.

Miller, M.R., Robinson, K.J., Cleary, M.D., Doe, C.Q., 2009. TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat. Methods 6, 439-441.

Molday, L.L., Wu, W.W.H., Molday, R.S., 2007. Retinoschisin (RS1), the protein encoded by the X-linked retinoschisis gene, is anchored to the surface of retinal photoreceptor and bipolar cells through its interactions with a Na/K ATPase-SARM1 complex. J. Biol. Chem. 282, 32792-32801.

Mori-Akiyama, Y., Akiyama, H., Rowitch, D.H., de Crombrugghe, B., 2003. Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. Proc. Natl. Acad. Sci. U.S.A. 100, 9360-9365.

Mori-Akiyama, Y., van den Born, M., van Es, J.H., Hamilton, S.R., Adams, H.P., Zhang, J., Clevers, H., de Crombrugghe, B., 2007. SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology 133, 539-546.

Morrison, S.J., Kimble, J., 2006. Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441, 1068-1074.

Mueller, T., Wullimann, M.F., 2002. Expression domains of neuroD (nrd) in the early postembryonic zebrafish brain. Brain Res. Bull. 57, 377-379.

Naya, F.J., Stellrecht, C.M., Tsai, M.J., 1995. Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. Genes Dev. 9, 1009-1019.

Nelson, S.B., Hempel, C., Sugino, K., 2006. Probing the transcriptome of neuronal cell types. Curr. Opin. Neurobiol. 16, 571-576.

Ng, L.J., Wheatley, S., Muscat, G.E., Conway-Campbell, J., Bowles, J., Wright, E., Bell, D.M., Tam, P.P., Cheah, K.S., Koopman, P., 1997. SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. Dev. Biol. 183, 108-121.

van Noort, V., Snel, B., Huynen, M.A., 2003. Predicting gene function by conserved co-expression. Trends Genet. 19, 238-242.

Nunes, M.C., Roy, N.S., Keyoung, H.M., Goodman, R.R., McKhann, G., 2nd, Jiang, L., Kang, J., Nedergaard, M., Goldman, S.A., 2003. Identification and isolation of multipotential neural progenitor cells from the subcortical white matter of the adult human brain. Nat. Med. 9, 439-447.

Oates, A.C., Pratt, S.J., Vail, B., Yan Y.L., Ho, R.K., Johnson, S.L., Postlethwait, J.H., Zon, L.I., 2001. The zebrafish klf gene family. Blood 98, 1792-1801.

Ochocinska, M.J. Hitchcock, P.F., 2007. Dynamic expression of the basic helix-loop-helix transcription factor neuroD in the rod and cone photoreceptor lineages in the retina of the embryonic and larval zebrafish. J. Comp. Neurol. 501, 1-12.

Ohshiro, T., Yagami, T., Zhang, C., Matsuzaki, F., 2000. Role of cortical tumour-suppressor proteins in asymmetric division of Drosophila neuroblast. Nature 408, 593-596.

Ohtoshi, A., Wang, S.W., Maeda, H., Saszik, S.M., Frishman, L J., Klein, W.H., Behringer, R.R., 2004. Regulation of retinal cone bipolar cell differentiation and photopic vision by the CVC homeobox gene Vsx1. Curr. Biol. 14, 530-536.

Ouyang, Y., Petritsch, C., Wen, H., Jan, L., Jan, Y.N., Lu, B., 2011. Dronc caspase exerts a non-apoptotic function to restrain phospho-Numb-induced ectopic neuroblast formation in Drosophila. Development 138, 2185-2196.

Pagliarini, R.A., Xu, T., 2003. A genetic screen in Drosophila for metastatic behavior. Science 302, 1227-1231.

Parmentier, M.L., Woods, D., Greig, S., Phan, P.G., Radovic, A., Bryant, P., O'Kane, C.J., 2000. Rapsynoid/partner of inscuteable controls asymmetric division of larval neuroblasts in Drosophila. J. Neurosci. 20, RC84.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L., et al., 2008. An integrated genomic analysis of human glioblastoma multiforme. Science 321, 1807-1812.

Passini, M.A., Kurtzman, A.L., Canger, A.K., Asch, W.S., Wray, G.A., Raymond, P.A., Schechter, N., 1998. Cloning of zebrafish vsx1: expression of a paired-like homeobox gene during CNS development. Dev. Genet. 23, 128-141.

Passini, M.A., Levine, E.M., Canger, A.K., Raymond, P.A., Schechter, N., 1997. Vsx-1 and Vsx-2: differential expression of two paired-like homeobox genes during zebrafish and goldfish retinogenesis. J. Comp. Neurol. 388, 495-505.

Peng, C.Y., Manning, L., Albertson, R., Doe, C.Q., 2000. The tumour-suppressor genes lgl and dlg regulate basal protein targeting in Drosophila neuroblasts. Nature 408, 596-600.

Perrière, G., Gouy, M., 1996. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78, 364-369.

Persson, A.I., Petritsch, C., Swartling, F.J., Itsara, M., Sim, F.J., Auvergne, R., Goldenberg, D.D., Vandenberg, S.R., Nguyen, K.N., Yakovenko, S., et al., 2010. Non-stem cell origin for oligodendroglioma. Cancer Cell 18, 669-682.

Petersen, P.H., Zou, K., Hwang, J.K., Jan, Y.N., Zhong, W., 2002. Progenitor cell maintenance requires numb and numblike during mouse neurogenesis. Nature 419, 929-934.

Piotrowski, T., Schilling, T.F., Brand, M., Jiang, Y.J., Heisenberg, C.P., Beuchle, D., Grandel, H., van Eeden, F.J., Furutani-Seiki, M., Granato, M., et al., 1996. Jaw and branchial arch mutants in zebrafish II: anterior arches and cartilage differentiation. Development 123, 345-356.

Pollard, S.M., Conti, L., Sun, Y., Goffredo, D., Smith, A., 2006. Adherent neural stem (NS) cells from fetal and adult forebrain. Cereb. Cortex 16 Suppl 1, i112-120.

Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., et al., 1998. Vertebrate genome evolution and the zebrafish gene map. Nat. Genet. 18, 345-349.

Postlethwait, J.H., Amores, A., Cresko, W., Singer, A., Yan, Y.L., 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 20, 481-490.

Potter, C.J., Huang, H., Xu, T., 2001. Drosophila Tsc1 functions with Tsc2 to antagonize insulin signaling in regulating cell growth, cell proliferation, and organ size. Cell 105, 357-368.

Rauch, G.J., Lyons, D.A., Middendorf, I., Friedlander, B., Arana, N., Reyes, T., Talbot, W.S., 2003. Submission and Curation of Gene Expression Data.

Rawls, J.F., Samuel, B.S., Gordon, J.I., 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. Proc. Natl. Acad. Sci. U.S.A. 101, 4596-4601.

Reilly, K.M., Loisel, D.A., Bronson, R.T., McLaughlin, M.E., Jacks, T., 2000. Nf1;Trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. Nat. Genet. 26, 109-113.

Rinn, J.L., Wang, J.K., Allen, N., Brugmann, S.A., Mikels, A.J., Liu, H., Ridky, T.W., Stadler, H.S., Nusse, R., Helms, J.A., et al., 2008. A dermal HOX transcriptional program regulates site-specific epidermal fate. Genes Dev. 22, 303-307.

Robinow, S., White, K., 1988. The locus elav of Drosophila melanogaster is expressed in neurons at all developmental stages. Dev. Biol. 126, 294-303.

Rodríguez-Marí, A., Yan, Y.L., Bremiller, R.A., Wilson, C., Cañestro, C., Postlethwait, J.H., 2005. Characterization and expression pattern of zebrafish Anti-Müllerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. Gene Expr. Patterns 5, 655-667.

Rolls, M.M., Albertson, R., Shih, H.P., Lee, C.Y., Doe, C.Q., 2003. Drosophila aPKC regulates cell polarity and cell proliferation in neuroblasts and epithelia. J. Cell Biol. 163, 1089-1098.

Roy, P.J., Stuart, J.M., Lund, J., Kim, S.K., 2002. Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature 418, 975-979.

Sabherwal, N., Tsutsui, A., Hodge, S., Wei, J., Chalmers, A.D., Papalopulu, N., 2009. The apicobasal polarity kinase aPKC functions as a nuclear determinant and regulates cell proliferation and fate during Xenopus primary neurogenesis. Development 136, 2767-2777.

Saint-Germain, N., Lee, Y.H., Zhang, Y., Sargent, T.D., Saint-Jeannet, J.P., 2004. Specification of the otic placode depends on Sox9 function in Xenopus. Development 131, 1755-1763.

Saito, T., Ikeda, T., Nakamura, K., Chung, U., Kawaguchi, H., 2007. S100A1 and S100B, transcriptional targets of SOX trio, inhibit terminal differentiation of chondrocytes. EMBO Rep. 8, 504-509.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

Sakai, D., Suzuki, T., Osumi, N., Wakamatsu, Y., 2006. Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development. Development 133, 1323-1333.

Sandell, J.H., Martin, S.C., Heinrich, G., 1994. The development of GABA immunoreactivity in the retina of the zebrafish (Brachydanio rerio). J. Comp. Neurol. 345, 596-601.

De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B., Sudbeck, P., Scherer, G., Poulat, F., Berta, P., 1998. Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. Mol. Cell. Biol. 18, 6653-6665.

Sauer, C.G., Gehrig, A., Warneke-Wittstock, R., Marquardt, A., Ewing, C.C., Gibson, A., Lorenz, B., Jurklies, B., Weber, B.H., 1997. Positional cloning of the gene associated with X-linked juvenile retinoschisis. Nat. Genet. 17, 164-170.

Schadt, E.E., Monks, S.A., Friend, S.H., 2003. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. Biochem. Soc. Trans. 31, 437-443.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467-470.

Schilham, M.W., Oosterwegel, M.A., Moerer, P., Ya, J., de Boer, P.A., van de Wetering, M., Verbeek, S., Lamers, W.H., Kruisbeek, A.M., Cumano, A., et al., 1996. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature 380, 711-714.

Schüller, U., Heine, V.M., Mao, J., Kho, A.T., Dillon, A.K., Han, Y.G., Huillard, E., Sun, T., Ligon, A.H., Qian, Y., et al., 2008. Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. Cancer Cell 14, 123-134.

Seimiya, M., Gehring, W.J., 2000. The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. Development 127, 1879-1886.

Sekiya, I., Tsuji, K., Koopman, P., Watanabe, H., Yamada, Y., Shinomiya, K., Nifuji, A., Noda, M., 2000. SOX9 enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6. J. Biol. Chem. 275, 10738-10744.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., Sander, M., 2007. SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. U.S.A. 104, 1865-1870.

Shen, Y., Raymond, P.A., 2004. Zebrafish cone-rod (crx) homeobox gene promotes retinogenesis. Dev. Biol. 269, 237-251.

Singh, S.K., Hawkins, C., Clarke, I.D., Squire, J.A., Bayani, J., Hide, T., Henkelman, R.M., Cusimano, M.D., Dirks, P.B., 2004. Identification of human brain tumour initiating cells. Nature 432, 396-401.

Southall, T.D., Brand, A.H., 2009. Neural stem cell transcriptional networks highlight genes essential for nervous system development. EMBO J. 28, 3799-3807.

Spana, E.P., Doe, C.Q., 1995. The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in Drosophila. Development 121, 3187-3195.

Spokony, R.F., Aoki, Y., Saint-Germain, N., Magner-Fink, E., Saint-Jeannet, J.P., 2002. The transcription factor Sox9 is required for cranial neural crest development in Xenopus. Development 129, 421-432.

Staugaitis, S.M., Zerlin, M., Hawkes, R., Levine, J.M., Goldman, J.E., 2001. Aldolase C/zebrin II expression in the neonatal rat forebrain reveals cellular heterogeneity within the subventricular zone and early astrocyte differentiation. J. Neurosci. 21, 6195-6205.

Stenkamp, D.L., 2007. Neurogenesis in the fish retina. Int. Rev. Cytol. 259, 173-224.

Stolt, C.C., Lommes, P., Sock, E., Chaboissier, M.C., Schedl, A., Wegner, M., 2003. The Sox9 transcription factor determines glial fate choice in the developing spinal cord. Genes Dev. 17, 1677-1689.

Sugimoto, M., Kimura, T., Tsumaki, N., Matsui, Y., Nakata, K., Kawahata, H., Yasui, N., Kitamura, Y., Nomura, S., Ochi, T., 1998. Differential in situ expression of alpha2(XI) collagen mRNA isoforms in the developing mouse. Cell Tissue Res. 292, 325-332.

Südbeck, P., Schmitz, M.L., Baeuerle, P.A., Scherer, G., 1996. Sex reversal by loss of the C-terminal transactivation domain of human SOX9. Nat. Genet. 13, 230-232.

Tanke, H.J., van der Keur, M., 1993. Selection of defined cell types by flow-cytometric cell sorting. Trends Biotechnol. 11, 55-62.

Tapon, N., Ito, N., Dickson, B.J., Treisman, J.E., Hariharan, I.K., 2001. The Drosophila tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. Cell 105, 345-355.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., Van de Peer, Y., 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. 13, 382-390.

Thisse, B., Thisse, C., 2004. Fast release clones: a high throughput expression analysis.

Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q., et al., 2001. Expression of the zebrafish genome during embryogenesis.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., Rubin, G.M., 2007. Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. 8, R145.

Toy, J., Yang, J.M., Leppert, G.S., Sundin, O.H., 1998. The optx2 homeobox gene is expressed in early precursors of the eye and activates retina-specific genes. Proc. Natl. Acad. Sci. U.S.A. 95, 10643-10648.

Uv, A.E., Harrison, E.J., Bray, S.J., 1997. Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. Mol. Cell. Biol. 17, 6727-6735.

Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al., 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17, 98-110.

Vidal, V.P., Chaboissier, M.C., de Rooij, D.G., Schedl, A., 2001. Sox9 induces testis development in XX transgenic mice. Nat. Genet. 28, 216-217.

Viktorin, G., Riebli, N., Popkova, A., Giangrande, A., Reichert, H., 2011. Multipotent neural stem cells generate glial cells of the central complex through transit amplifying intermediate progenitors in Drosophila brain development. Dev. Biol. 356, 553-565.

Visel, A., Thaller, C., Eichele, G., 2004. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res. 32, D552-556.

Visvader, J.E., 2011. Cells of origin in cancer. Nature 469, 314-322.

Vogel, W., 1999. Discoidin domain receptors: structural relations and functional implications. FASEB J. 13 Suppl, S77-82.

Vogel, W.F., Abdulhussein, R., Ford, C.E., 2006. Sensing extracellular matrix: an update on discoidin domain receptor function. Cell. Signal. 18, 1108-1116.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., et al., 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 79, 1111-1120.

Wallace, K., Liu, T.H., Vaessin, H., 2000. The pan-neural bHLH proteins DEADPAN and ASENSE regulate mitotic activity and cdk inhibitor dacapo expression in the Drosophila larval optic lobes. Genesis 26, 77-85.

Wang, H., Somers, G.W., Bashirullah, A., Heberlein, U., Yu, F., Chia, W., 2006. Aurora-A acts as a tumor suppressor and regulates self-renewal of Drosophila neuroblasts. Genes Dev. 20, 3453-3463.

Wang, Y., Yang, J., Zheng, H., Tomasek, G.J., Zhang, P., McKeever, P.E., Lee, E.Y.H.P., Zhu, Y., 2009. Expression of mutant p53 proteins implicates a lineage relationship between neural stem cells and malignant astrocytic glioma in a murine model. Cancer Cell 15, 514-526.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57-63.

Weng, M., Golden, K.L., Lee, C.Y., 2010. dFezf/Earmuff maintains the restricted developmental potential of intermediate neural progenitors in Drosophila. Dev. Cell 18, 126-135.

Westerfield, M., 2000. The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio) University of Oregon Press, Eugene, OR.

Wilson, M.E., Yang, K.Y., Kalousova, A., Lau, J., Kosaka, Y., Lynn, F.C., Wang, J., Mrejen, C., Episkopou, V., Clevers, H.C., et al., 2005. The HMG box transcription factor Sox4 contributes to the development of the endocrine pancreas. Diabetes 54, 3402-3409.

Wu, W.W.H., Wong, J.P., Kast, J., Molday, R.S., 2005. RS1, a discoidin domain-containing retinal cell adhesion protein associated with X-linked retinoschisis, exists as a novel disulfide-linked octamer. J. Biol. Chem. 280, 10721-10730.

Xiong, W.C., Okano, H., Patel, N.H., Blendy, J.A., Montell, C., 1994. repo encodes a glial-specific homeo domain protein required in the Drosophila nervous system. Genes Dev. 8, 981-994.

Ya, J., Schilham, M.W., de Boer, P.A., Moorman, A.F., Clevers, H., Lamers, W.H., 1998. Sox4-deficiency syndrome in mice is an animal model for common trunk. Circ. Res. 83, 986-994.

Yan, Y.L., Hatta, K., Riggleman, B., Postlethwait, J.H., 1995. Expression of a type II collagen gene in the zebrafish embryonic axis. Dev. Dyn. 203, 363-376.

Yan, Y.L., Miller, C.T., Nissen, R.M., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P.A., Amsterdam, A., et al., 2002. A zebrafish sox9 gene required for cartilage morphogenesis. Development 129, 5065-5079.

Yan, Y.L., Willoughby, J., Liu, D., Crump, J.G., Wilson, C., Miller, C.T., Singer, A., Kimmel, C., Westerfield, M., Postlethwait, J.H., 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132, 1069-1083.

Yang, Z.J., Ellis, T., Markant, S.L., Read, T.A., Kessler, J.D., Bourboulas, M., Schüller, U., Machold, R., Fishell, G., Rowitch, D.H., et al., 2008. Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. Cancer Cell 14, 135-145.

Zeiner, G.M., Cleary, M.D., Fouts, A.E., Meiring, C.D., Mocarski, E.S., Boothroyd, J.C., 2008. RNA analysis by biosynthetic tagging using 4-thiouracil and uracil phosphoribosyltransferase. Methods Mol. Biol. 419, 135-146.

Zhang, P., Jimenez, S.A., Stokes, D.G., 2003. Regulation of human COL9A1 gene expression. Activation of the proximal promoter region by SOX9. J. Biol. Chem. 278, 117-123.

Zhu, X., Bergles, D.E., Nishiyama, A., 2008. NG2 cells generate both oligodendrocytes and gray matter astrocytes. Development 135, 145-157.

Zhu, Y., Romero, M.I., Ghosh, P., Ye, Z., Charnay, P., Rushing, E.J., Marth, J.D., Parada, L.F., 2001. Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. Genes Dev. 15, 859-876.

Zhu, Y., Guignard, F., Zhao, D., Liu, L., Burns, D.K., Mason, R.P., Messing, A., Parada, L.F., 2005. Early inactivation of p53 tumor suppressor gene cooperating with NF1 loss induces malignant astrocytoma. Cancer Cell 8, 119-130.

Zhuo, L., Theis, M., Alvarez-Maya, I., Brenner, M., Willecke, K., Messing, A., 2001. hGFAP-cre transgenic mice for manipulation of glial and neuronal function in vivo. Genesis 31, 85-94.

Zong, H., Espinosa, J.S., Su, H.H., Muzumdar, M.D., Luo, L., 2005. Mosaic analysis with double markers in mice. Cell 121, 479-492.